Investigating the age differences in good enough processing: eye-tracking evidence from young and older adults

A thesis submitted for the degree of

Doctor of Philosophy

at the University of Nottingham

by

Taminder Kaur Kalsi

June 2022

Abstract

Shallow processing and good enough processing are two theoretical accounts which advocate that a full syntactic and semantic analysis may not always be accomplished, instead readers often rely on heuristics, such as context and word knowledge (Ferreira et al., 2002; Sanford & Graesser, 2006; Sanford & Sturt, 2002). This type of processing often leads to readers establishing a shallow, underspecified or 'good enough' mental representation of text. Good enough processing has been widely investigated in a young adult population. Another particularly interesting population to explore good enough processing in are healthy older adults because across the lifespan, elements of crystallised intelligence, such as world knowledge and vocabulary expand (Paterson et al., 2020; Salthouse, 2009). However, elements of fluid intelligence, for instance, memory can deteriorate across the lifespan (Daneman & Merikle, 1996; Salthouse & Babcock, 1991). Coupled together, older readers may be more inclined to rely on heuristics, such as world knowledge than their younger counterparts to compensate for deterioration of elements of fluid intelligence (Christianson et al., 2006). This thesis comprises of four eye-tracking experiments which examine the online processing of young and older adults whilst reading sentences that could bias them to a good enough interpretation at the word-, sentence- and wider discourse-level. In Experiment 1, sentences containing hard-to-detect anomalies were presented to assess shallow processing at the word-level. In Experiment 2, sentences containing temporary syntactic ambiguities were presented and in Experiment 3 plausible and implausible sentences in active or passive sentence structure were used to assess good enough processing at the sentence-level. Finally, doubly quantified sentences can create semantic ambiguity and this was exploited to examine the discourse-level. There were three key findings from this thesis. Firstly, in contrast to the prediction of the author, older adults were no more likely to engage in good enough processing than younger readers (Experiments 2 and 4). Instead, in Experiments 1 and 3, older adults had higher accuracy scores (which were used as an index of good enough processing), therefore, suggesting they engaged in good enough processing less than their younger counterparts. Eye tracking results also revealed that young and older adults may use different strategies during online processing of ambiguities but ultimately achieve similar levels of comprehension (Experiments 2 and 3). The relationship between eye movements made during online processing of text and comprehension accuracy was examined in Experiment 1, 2 and 3. Another key finding was that various eye movements were found to predict good enough processing, especially regressions. Generally an increase in regressions was identified as a predictor of less good enough processing in all three experiments. The conclusion from this thesis is that whilst shallow and good enough processing accounts can account for some of the findings, results from Experiments 1, 3 and 4 suggest that these frameworks cannot account for all of the findings.

Acknowledgements

I would not have completed this thesis if it were not for the support of many people.

First and foremost, thank you to my supervisors Dr Ruth Filik and Professor Kevin Paterson for selecting me for this PhD programme and for all of the support and guidance along the way. I am extremely grateful for their advice, their time and for always encouraging me to be better. I feel very privileged to have had them as my supervisors and I have learned so much from them.

I would like to thank the Cognition and Language group at University of Nottingham and the Vision and Language group at University of Leicester. Special thanks to Sammie Tyler for co-leading the ReproducibiliTEA group at University of Leicester with me. I would also like to thank Dr Kayleigh Warrington and Dr Vicky McGowan for the insightful conversations. A particular thank you to Dr Ascen Pagan for being a pillar of support over the past few years.

I would like to thank the younger and older adult participants in the East Midlands area. I conducted these experiments on some amazing people!

Thank you to the ESRC for funding my PhD programme and awarding me a three-month policy placement at the Department for Education, which quite possibly changed the course of my career. Also, thank you to the Experimental Psychology Society for invaluable travel grants which enabled me to attend and present at many interesting conferences.

On a personal note, I want to thank my support network; my friends and family – Jagdip, Manpreet, David, Anaiya, Ariya and Anya for their encouragement and kindness during these last few years. A particular thanks to Sarabjit and Benjamin for their unwavering support and unconditional love throughout, I could not have done this without you.

Declaration

I, Taminder Kalsi, declare that the work described in this thesis is my own and was carried out during my PhD programme. Furthermore, all relevant materials and sources have been cited accordingly.

Contents

List of tables

List of figures

List of abbreviations

ANOVA = Analysis of Variance

ERP = Event Related Potential

GLME = General Linear Mixed Effects Models

LMEM = Linear Mixed Effects Models

M = Mean

MoCA = Montreal Cognitive Assessment

SE = Standard Error

SD = Standard Deviation

WAIS-IV = Weschler Adult Intelligence Scale – fourth edition

WM = Working Memory

Chapter 1: Reading and Aging

1.1. Background

An accurate representation of text is crucial for one to operate in society to meet typical tasks required of an individual in their working and personal life, such as reading application forms, reports, news articles and announcements. Sentence comprehension is often thought to be a smooth process where words are accurately assembled into a syntactic structure producing a coherent representation (Dede & Flaxx, 2016; Starr & Rayner, 2001). Whilst reading might seem effortless, it is actually a highly complex process, one that requires the orchestration of most of the major faculties in the brain (such as, attention, vision, memory and eye movement control) in order to build a complete and accurate mental interpretation of text (Price, 2012). Two frameworks, good enough processing (Ferreira et al., 2002) and shallow processing (Sanford & Sturt, 2002) advocate readers often utilise *fast and frugal* heuristics relying on world knowledge and context to derive an interpretation rather than fully completing syntactic and/or semantic processing. Consequently, sentence comprehension results in a partial, superficial, or underspecified representation. Numerous empirical studies have found older adults experience greater difficulty during reading (Kemper & McDowd, 2006; Kliegl et al., 2004; McGowan et al., 2015; McGowan et al., 2014; Paterson et al., 2013a, 2013b; Rayner et al., 2006, 2011, 2013; Whitford & Titone, 2017, 2016). Across the lifespan, more world knowledge is accumulated and so older adults may compensate for any cognitive decline and difficulties reading by employing shallow or good enough processing.

This thesis explores age differences in shallow and good enough processing at the word-, sentence- and discourse-level for young and older adults. Chapter 1 focuses on a brief introduction to aging and reading. Chapter 2 explains the shallow and good enough processing frameworks and key factors affecting depth of processing. Chapter 3 describes the methods employed in this thesis. The experimental studies can be found in Chapters 4 – 7. Finally, Chapter 8 is the General Discussion.

1.2. Eye movements during reading in healthy adults

Reading requires the reader moving their eyes in a highly systematic way by making a sequence of rapid eye movements (saccades) that carry your gaze from word to word, separated by relatively brief pauses (fixations). The average length of a saccade is around 7 – 9 letters and an average fixation lasts between 200 – 250ms in healthy young adult readers during which your eyes take in information from the page (Liversedge & Finlay, 2000). The fovea is a tiny area covering no more than 1 ½ millimetres located within the retina, which contains the highest density of cells needed to see a word in the sharpest detail (Liversedge & Finlay, 2000). In this way, we read by seeing a series of small chunks of information that the brain must connect together to make sense of the text as a whole. Forward saccades account for 85% of eye movements when reading, however, around 15% of the time, readers' eye movements move back to earlier portions of the text, usually to fix errors and uncertainties, or to help connect information from different parts of the text (Liversedge & Finlay, 2000; Staub & Rayner, 2007). Somewhat surprisingly, we skip some words altogether during reading. The age that words are acquired, as well as the frequency, predictability and length of words can often influence whether words are skipped by the reader (Balota et al., 2006; Duffy et al., 1988; Staub & Rayner, 2007).

In contrast to healthy young adult readers, healthy older adults have been found to have longer reading times suggesting that they encounter greater difficulty when reading (Kemper & McDowd, 2006; Kliegl et al., 2004; McGowan et al., 2015; McGowan et al., 2014; Paterson et al., 2013a, 2013b; Rayner et al., 2006, 2011, 2013; Whitford & Titone, 2017, 2016) especially with long and complex sentences (Obler et al., 1991). Despite this, there is some evidence that offline comprehension remains preserved into late adulthood (Artuso & Belacchi, 2021; Baltes et al., 1999; Christianson et al., 2006; De Beni et al., 2003).

The remainder of this chapter explains some of the visual factors that change across the lifespan (**1.2**), age differences in online processing of text between young and older adults (**1.3**), key cognitive factors that may affect reading (**1.4**) and finally strategies older adults may use to compensate for changes across the lifespan (**1.5**).

1.3. Visual factors that affect reading during aging

There are various visual factors that can affect reading in healthy older adults who have an absence of any oculomotor disease and have corrected to normal vision (Owsley, 2011). Visual decline that occurs naturally in healthy aging may impact older adults' language processing when reading text (Paterson et al., 2020).

Visual acuity generally refers to the sharpness of vision (Anderson, 2017). During healthy aging, there is a general decline in the ability to process visual information (Owsley, 2011). Some of the reasons for this decline may be due to reduced retinal illumination (Owsley, 2011) or difficulty in visual crowding (McGowan & Reichle, 2018) which is where a particular letter or word is difficult to process due to interference from external letters (see Figure 1).

Figure 1

An example of visual crowding is where 'A' is crowded by 'C' and 'T' on the right hand side making it more difficult to process compared to the 'A' on the left hand side

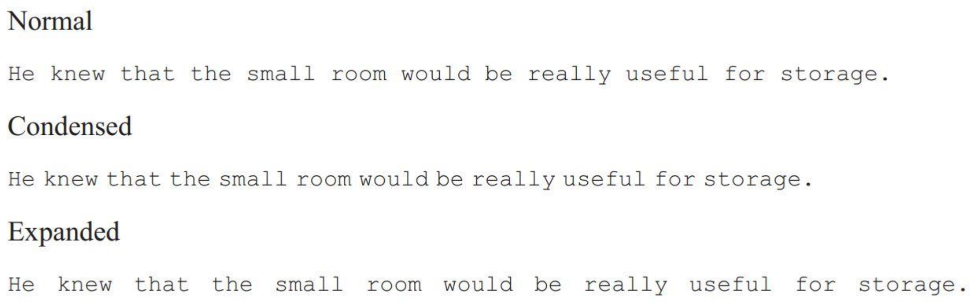
A + CAT

Evidence of this comes from Rayner et al. (2013) who conducted an eye-tracking study where young and older adults read spaced text (e.g., *Christine put the money in her account for safekeeping*) and text that contained no spaces after each word (e.g., *Christineputthemoneyinheraccountforsafekeeping*). Older adults experienced greater difficulty when reading text that contained no spaces after each word (Rayner et al., 2013). Overall, the findings from this experiment illustrate the difficulty for the eyes to move to specific locations (usually the preferred viewing position is just left of the centre of a word) and as a result this makes the identification of individual words more difficult (Rayner et al., 2013).

Another interesting study examined the effects of inter-word spacing by presenting text that was normally spaced, half the normal space or where the spacing was 1.5 times greater than typically presented text (see Figure 2) (McGowan et al., 2015). Findings from the study showed young and older adults performance was no different in inter-word spacing suggesting that older adults adjust their eye movement behaviour when text is subtly different as long as some spacing is available (McGowan et al., 2015). This study further supported the proposal that changes in reading performance during aging are not due to issues with eye movement control as older adults were able to adjust their eye movement behaviour when reading text (McGowan et al., 2015).

Figure 2

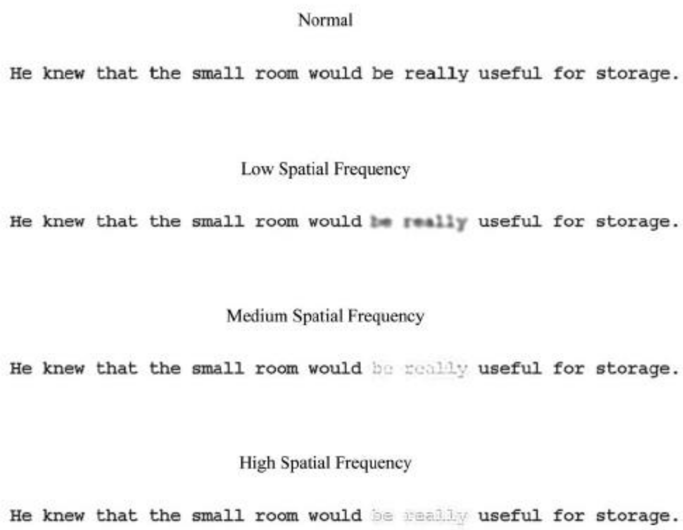
Example of inter-word spacing conditions in McGowan et al.'s (2015) study



Furthermore, contrast sensitivity has been found to deteriorate throughout adulthood (Hahn et al., 2009; Owsley et al.,1983). Contrast sensitivity is integral for processing visual information in fine detail. For instance, Jordan et al. (2014) used a filtering technique to display text that was presented in low, medium or high spatial frequencies as well as normal text which acted as a control (see Figure 3).

Figure 3

An example of sentences presented in normal, low, medium, and high spatial frequencies from the Jordan et al. (2014) study



High spatial frequencies provide fine visual detail and low spatial frequencies display coarse-grained detail. In this study, the findings indicated that when compared to medium and high spatial frequencies, younger adults experienced longer reading times and longer fixation durations for low spatial frequencies. Reading times for normal text and medium and high spatial frequencies in younger adults were no different. In contrast, older adults had longer reading times and longer fixation durations for all text in low, medium and high spatial frequencies compared to normally displayed text. Furthermore, older adults had faster reading times and shorter fixations durations for text presented in medium spatial frequencies than low and high spatial frequencies. The results indicated that young adults performed better when high visual detail was available from high spatial frequencies. It seems older adults have greatest difficulty with high spatial frequencies (McGowan et al., 2014; Paterson et al., 2013b). This study highlighted that it is not just high visual detail that is required for processing text during reading, instead text can be processed when presented in various different spatial frequencies. Furthermore, although differences in reading text at different spatial frequencies was found amongst both age groups, reading text remained preserved through to later adulthood.

To address the visual factors affecting reading in older adults, visual acuity and contrast sensitivity tests were administered to each participant in the studies detailed in this thesis. The most common visual acuity and contrast sensitivity tests are measured using Snellen and logMAR units (Anderson, 2017). Visual acuity was tested using either the Early Treatment Diabetic Retinopathy Study (ETDRS) (Ferris & Bailey, 1996) or the Sloan visual acuity chart. Contrast sensitivity was assessed using the Pelli–Robson chart (Pelli et al., 1988) or the Sloan contrast sensitivity chart. Differences in tests occurred due to testing in different labs. Visual acuity was tested with normal to corrected vision (if required) for younger and older adults. Data of participants with an acuity of 20/60 or above at a distance of 40 cm were discarded in line with clinical practice. No limit was applied to contrast sensitivity.

In addition to visual acuity and contrast sensitivity, reading acuity was assessed as this reflects the nature of reading (Chen et al., 2019; Legge et al., 1992) in that reading words and sentences results in greater visual complexity because individual letters are more difficult to identify within words and sentences (Radner, 2017). Furthermore, sentences contain capital and lower-case letters increasing the complexity of visual information an individual processes. Thus, it was important to assess reading acuity as these charts better reflect the nature of reading than the commonly presented single letter eye charts (also referred to as an optotype). In the experimental studies, participants also read the Radner reading chart for a comprehensive evaluation of visual acuity (Radner, 2017; Radner et al., 2002). Again, this chart was measured using Snellen units, no exclusion criteria were applied.

1.4. Language processing and aging

The above section detailed some of the important visual factors associated with processing text and how these change across the lifespan. Additionally, there are a range of linguistics factors affecting the processing of text at the word, sentence and wider discourse level.

1.4.1. Word level processing

Broadly speaking, word recognition is the combination of successful identification of orthographic and phonological representations which then initiates the semantic representation of the given word (see Balota et al., 2006 for a review). Common factors that affect word recognition are length, frequency and predictability of the word (Kliegl et al., 2004; Rayner, Reichle, et al., 2006). Furthermore, lexical neighbours can also affect word recognition (Warrington, Mcgowan, et al., 2018). The next two sections describe how these factors affect word recognition in further detail.

1.4.1.1. Length, frequency and predictability

Literature has found that age differences occur in word identification. Kliegl et al. (2004) used a corpus dataset and tested how the length, frequency and predictability of words affect processing in young (19-28 year olds) and older adults (65-83 year olds) by measuring inspection probability (skipping), first fixation, single fixation, gaze duration and total reading times. Older adults generally read more slowly and were more sensitive to frequency, meaning they responded faster when a word was of high frequency but more slowly compared to young adults when the word was of low frequency. There were qualitative age differences in predictability effects in that both groups were sensitive to predictability but older adults made fewer refixations and young adults were more likely to skip highly predictable words. There were few age differences found in word length effects.

Similarly, Rayner et al. (2006) examined the effects of word frequency, predictability and font type in young (18-34 years old) and older adults (70-92 years old) using eye-tracking. The experiment also manipulated the font text was presented in, the results showed that ‘Old English’ font created more difficulty than ‘Times New Roman’ for older than young adults, with older adults making more regressions than young adults. Results for frequency showed older adults were more sensitive to word frequency as longer single fixations, gaze durations and total reading times were found for this age group when reading low frequency words compared to young readers in the study. Findings from the experiment also showed that older adults had a higher likelihood of skipping words as well as landing further into words that were of a high frequency (e.g., *building*) compared to words of a low frequency (e.g., *catacomb*). However, the skipping rate for young adults was no different for high or low frequency words. This study, similarly to Kliegl et al. (2004), found predictability effects for both young and older adults who produced shorter fixations when words of a highly predictability were presented.

1.4.1.2. Lexical neighbours

How young and older adults perceive words has also been examined using lexical neighbours (Warrington et al., 2018). Words like ‘brunch’ and ‘branch’ are lexical neighbours as both have the same number of letters in the same order and only differ by one letter (u/a). Young adults have been found to commonly mistake high frequency neighbours e.g., ‘branch’ for ‘brunch’ (Slattery, 2009). Warrington et al. (2018) tasked both young (18-29 years) and older readers (65-79 years) with processing sentences which contained target words (e.g., *brunch)* that immediately followed a high frequency neighbour (e.g., *branch*) or did not have a high frequency lexical neighbour. Experimental items also included a bias manipulation with item either being biased or not biased by the sentence context (see Figure 4).

Figure 4

Example stimuli from Warrington et al.’s (2018) study

Context	Stimulus type	
Neutral context	Experimental	Due to the freezing rain, the *brunch* (branch) was postponed a week.
	Control	Due to the freezing rain, the *buffet* was postponed a week.
Biased context	Experimental	Everyone said that the food at the *brunch* (branch) was simply magnificent.
	Control	Everyone said that the food at the *buffet* was simply magnificent.

Target words are shown in italics and the HFN in parentheses.

Longer total reading times, rereading times and more regressions in to the target word were found for target words presented with a high frequency neighbour. Furthermore, longer total and rereading times were found for target words with a high frequency neighbour presented in neutral contexts than biased contexts. In accordance with previous research, young and older adults were commonly succeptible to lexical neighbour and sentence context effects and no age differences were identified. The findings from this study show that effects of lexical competition and context predictability remain preserved across the lifespan.

1.4.2. Sentence level processing

1.4.2.1. Context

At the sentence level, predictability is particularly interesting because eye-tracking studies and Event Related Potential (ERP) studies have found contrasting results. Eye-tracking studies have found that the semantic context of sentences (e.g., *Since the wedding was today, the baker rushed the wedding…*) makes the upcoming word highly predictable (i.e., *wedding cake* rather than *pie*) and this can facilitate comprehension in older adults more so than young adults (Balota et al., 1985). Furthermore, studies examining predictability in speech have supported eye-tracking studies (Sheldon et al., 2008). However, ERP studies have found a different pattern. Federmeier and Kutas (2005) presented sentences where it was easier to predict the last word (*beard*) when the context was highly constraining (e.g., *No one at the reunion recognized Dan because he had grown a beard)*. However, predictability was more difficult when the context was weakly constraining (e.g., *At the children's park next to the beach she saw a man with a beard*) (Federmeier & Kutas, 2005).

Federmeier and Kutas (2005) measured the electrical activity produced by the brain when reading sentences in young (18-24 years) and older adults (60-76 years). In weakly constraining contexts, both age groups showed an N400 magnitude, this is a component sensitive to semantics and context. In highly constraining contexts, older adults' processing was delayed and young adults exhibited a larger N400 effect suggesting that older reader may be inclined to use semantic context to predict words less than young readers. Overall, these results showed that in ERP studies older adults find it more difficult to utilise sentential context (Wlotko et al., 2010). However, these results may reflect differences in tasks as eye-tracking studies present whole sentences and ERP studies use Rapid Serial Visual Presentation (RSVP). Task requirements may be a factor as previous literature suggests that older readers may be more inclined to use context in tasks where reading time can be determined by the individual (Janse et al., 2007).

1.4.2.2. Verb type

Another factor that affects sentence processing is verb type. Christianson et al. (2006) presented sentences containing temporary syntactic ambiguities (e.g., 2a and 2b) with verbs that were either optionally transitive (e.g., *read* and *hunt*) or reflexive absolute transitive (e.g., *shave* and *dress*) (Experiment 1).

1a) While Anna dressed the baby played in the crib. (Ambiguous sentence, RAT verb)

1b While the man hunted the deer ran into the woods. (Ambiguous sentence, OPT verb)

Verbs that are reflexive absolute transitive in nature are unique in that when no object is included, the argument remains intact. For instance, 'Anna dressed*'* and 'Anna dressed herself' have identical meanings. Optionally transitive verbs (OPT) can have a transitive structure (e.g., *The man is hunting the deer*.) or an intransitive structure (*The man was hunting*.) where there is no specified object. Results showed that compared to young adult readers, older readers were more inclined to hold on to the transitive meaning of OPT verbs in ambiguous and unambiguous sentences (Christianson et al., 2006). There were no age differences for RAT verbs. In the context of the experimental sentences in the study, 'Anna dressed herself' and 'Anna dressed the baby' (1a) have different representations. However, ‘The man hunted [something]’ and ‘The man hunted the deer’ (1b) are compatible. Christianson et al. (2006) concluded that older readers are less likely to resolve ambiguities because this group are more likely to rely on gist and make inferences based on their knowledge and experience of the world to compensate for working memory decline across the lifespan.

Previous research has shown that memory traces of syntactic structures decay quickly (Sachs, 1967). Thus, the gist of the OPT verb in sentence (1b) is that the man is hunting and based on gist and world knowledge, it is conceivable that the man is hunting the deer and so older adults are more likely to answer ‘yes’ to comprehension questions like ‘*Was the man hunting the deer?’*. These findings have been replicated in sentence picture matching tasks (Malyutina & den Ouden, 2016), self-paced reading experiments (Kemtes & Kemper, 1997) and eye-tracking studies (Kemper et al., 2004). This example illustrates that verb type can affect sentence processing for young and especially for older adults.

1.4.2.3. Sentence structure

Another common finding is that older adults are more likely to experience greater difficulty processing complex sentence structures than their younger counterparts. For example, adults (30-79 year olds) listened to six types of sentence structures including, passives (2a), actives (2b), single negative (2c), double negative (2d), doubly embedded (2e) and comparatives (2f) (Obler et al., 1991).

2a) The fierce wolf attacked the lost sheep in the woods. (Active)

2b) The escaped convict was tracked by the sniffing hound through the forest. (Passive)

2c) The comment that was not appropriate irritated the secretary. (Single negative)

2d) The bureaucrat who was not dishonest refused the bible. (Double negative)

2e) The doctor who helped the patient who was sick was healthy. (Double embedded)

2f) The horsehair rug was rougher than the baby blanket. (Comparative)

Results showed that answers to questions required longer processing time for syntactically complex sentences in older adults than young adults (Obler et al., 1991). Age differences were also identified in accuracy rates. Older adults had significantly lower accuracy rates on implausible sentence suggesting that older adults are particularly vulnerable to errors because this group are more inclined to rely on world knowledge compared to young adults. Furthermore, older adults answered more questions incorrectly when sentences contained syntactic complexity compared to young adults. The authors suggested that for simple sentences, such as, active sentences (2a) there is a gradual decline in processing and for complex sentences, older adults are more likely to make more errors when reading complex sentences containing.

Interestingly, no differences in production difficulty across sentence structures have previously been identified. Altmann and Kemper (2006) presented three words on a screen (see Figure 5) to young (M_{age} = 20) and older adults (M_{age} = 76). Participants were required to create a sentence from the three words displayed, which always consisted of one verb and two nouns. Responses were scored as accurate if the sentence produced was grammatical, included all the words displayed on the screen and were fluent. Response time comprised of the onset of the stimuli until the participants began a response.

Figure 5

Example stimuli presented in Altmann and Kemper's (2006) sentence production experiment butler

stirred

juice

Overall, the results showed neither age group had preferences for producing sentence structures being active or passive. Participants were instead influenced by the order of the stimuli presented (animate object/inanimate object presented first), especially older adult participants. The results showed that in some conditions older adults even produced more passive sentence structures (although there was no difference overall). Furthermore, the hypothesis of cognitive aging which suggests there is a general slowing across the lifespan was not supported as few age differences occurred in processing time or accuracy. The authors instead suggested that few age differences occurred because of different strategies used by young and older adults. This is not the first study to suggest that young and older adults use different strategies (Kliegl et al., 2004; Stine-Morrow et al., 2001) and this is discussed further in section 1.5.

1.4.2.4. Plausibility

Another area where age differences have not been found to occur is in the processing of sentences containing plausibility violations. For instance, Yoon et al. (2015) presented semantically plausible and implausible sentences auditorily to young (18-35 years) and older adults (aged 55-75). In addition to control sentences (3a and 3b), sentences with one negative clause (3c and 3d) or two negative clauses (3e and 3f) were also presented to participants.

3a) Because the ceiling light is off, the room is dark. (Control, plausible)

3b) Because the ceiling light is on, the room is dark. (Control, implausible)

3c) Because the ceiling light is **not** on, the room is dark. (One negative, plausible)

3d) Because the ceiling light is **not** off, the room is dark. (One negative, implausible)

3e) Because the ceiling light is **not** off, the room is **not** dark. (Two negative, plausible)

3f) Because the ceiling light is **not** on, the room is **not** dark. (Two negative, implausible)

Participants were tasked with rating the sentences (3a-3f) presented as *likely* or *unlikely*. Findings showed that both young and older participants had lower accuracy rates for implausible sentences. Furthermore, an interaction showed that older adults were more sensitive to negative sentences with less accurate answers for sentences where there were two negative clauses compared to one. On the other hand, young adults had lower accuracy rates for sentences containing two negative clauses compared to control sentences. This study also examined executive function by testing working memory, inhibitory control and attention switching. An interaction between age and inhibitory control was identified showing that older adults with high inhibitory control had a similar performance to young adults. In contrast, older adults with low inhibitory control performed worse than young adults with low inhibitory control. The authors concluded that age differences may instead reflect differences in executive function, further detail on cognition and aging can be found in section 1.4.

Not all research points to a strong decrease in performance during aging as further evidence of preserved language processing across the lifespan has been found. Hardy et al. (2017) used a dialogue task to investigate age differences in syntactic priming. Participants heard active and passive sentences and then described target pictures. Both young (18-23 years old) and older adults (69-80 years old) were more likely to produce a description of the target picture in a passive structure after listening to passive sentences. This priming effect did not occur with active sentences. Furthermore, a lexical boost effect was found where the priming effect increased in instances where the prime and the target verb were the same.

These findings show that certain semantic and syntactic representations may be preserved across the lifespan.

1.4.3. Discourse level processing

At the discourse level of processing, studies have shown that older adults outperform young adults in text recall (Morrow et al., 1997; Radvansky et al., 1990). Radvansky et al. (2001) presented pieces of text to young (18-26 years) and older adults (aged 61-96) that focused on unfamiliar historical events (e.g., the British Gunpowder plots of 1605) which were presented across 58-85 lines. These studies are based on a design created by Schmalhofer and Glavanov (1986) where extracts of text are presented to participants who complete a recognition task after reading the texts. The recognition test included four different probes; a) verbatim probes directly from the text, b) paraphrased probes that contained the same arguments but were expressed differently, c) inference probes which were probes not directly mentioned but were consistent with the situation displayed and d) incorrect probes. Reading times were measured as well as recognition based on surface level, text base level knowledge and situation model level information. Surface level encompasses word recognition and syntax in the text. Text base level knowledge refers to understanding the gist of the exact text but not necessarily the exact details. Surface level knowledge relates to the events included in the text and an understanding of what the text is about. Results showed that young adults made fewer errors with surface-level and text base information compared to older adults. In contrast, older adults had fewer errors or no difference with situation based knowledge. The authors concluded young adults have superiority in knowledge regarding the details of the text, however the wider discourse remains preserved into later adulthood (Radvansky et al., 2003; Radvansky & Dijkstra, 2003; Radvansky et al., 2001).

Additionally, older adults have been found to have faster reading times when the discourse has a highly constraining context. Stine-Morrow et al., (1996) presented expository texts to young (M_{age} = 20) and older (M_{age} = 69) adults and found older adults better utilised context compared to younger adults which was related to a higher recall. This reflects a facilitation effect that occurs due to an ability to utilise context more so than if the context was not specific.

1.4.3.1. Anaphora

An interesting phenomenon to study that links language and discourse is anaphora. Successful anaphor resolution requires the reader to link the antecedent (e.g., *firefighter*) to the anaphor pronoun (e.g., *himself*) when encountering it downstream in the sentence (e.g., *The firefighter burned himself while rescuing victims from the building*). In three eye-tracking experiments, Shake and Stine-Morrow (2011) created sentences with anaphor violations by presenting sentences with occupations that are traditionally associated with a specific gender (e.g., *The firefighter burned herself while rescuing victims from the building*). Results showed that older adults were more likely to identify violations when sentences were embedded in a paragraph of text (Experiment 2a and 2b) compared to single line texts (Experiment 1). However, younger adults were much quicker at resolving anaphors than older adults as shorter first fixations and reprocessing time was found for young adults compared to older adults.

In summary, whether the evidence regarding age differences exist in reading seems somewhat mixed. A breadth of research shows that age differences are apparent. Despite older adults experiencing more difficulty in processing text, reading comprehension often remains preserved across the lifespan. A number of strategies that older adults may use to compensate for reading difficulty have been found and are described in section (1.5).

1.5. Aging and strategies during reading

Despite older adults experiencing greater reading difficulty when processing text, there is some evidence that suggests generally, comprehension remains preserved (Artuso & Belacchi, 2021; Baltes et al., 1999; De Beni et al., 2003). To account for this difference, it has been suggested that young and older adults use different strategies during language processing. This section describes some of the strategies that have been put forward.

1.5.1. Stimulus characteristics

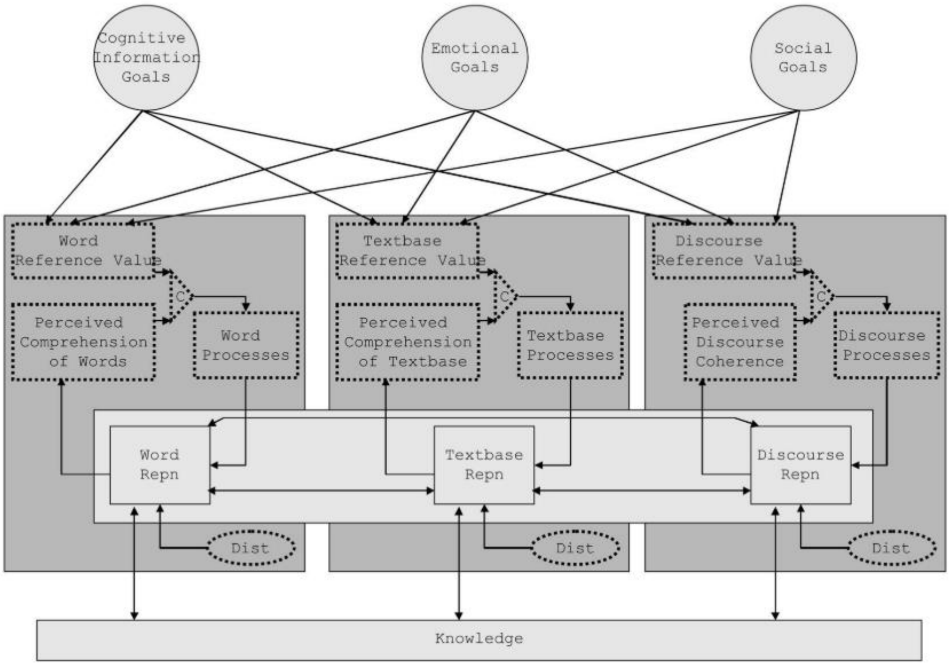
As mentioned previously, Altmann and Kemper (2006) presented young and older adults with three words (see Figure 5) and participants were asked to create a sentence. Despite theories suggesting that older adults would create more active than passive sentences to minimise the high cognitive load that complex sentences require, no such pattern was found and in some conditions, older adults created more passive sentences than their younger counterparts. As a result, the hypothesis of cognitive aging and the inhibitory deficit hypothesis could not account for the findings. Instead, Altmann and Kemper (2006) suggested that young and older adults used different strategies, specifically, older adults rely on their environment, in this instance, the order of the stimuli presented. This is reflective of the finding that older adults were more likely to create a sentence structure based on the noun-verb-noun order displayed. For instance, in inanimate-object conditions, older adults created more inanimate-object sentences than all other conditions. This may reflect the reduced resources theory of aging (Craik, 1986; Scheffers et al., 1999) where older adults compensate for cognitive decline by having a greater reliance on stimuli characteristics. In contrast to older adults, young adults displayed a more cautious strategy where they produced sentences once having reviewed all the available stimuli on the screen and then choosing the preferred structure. The time taken to review all the stimuli on the screen may have contributed to the slow reaction times found for this group, this may account for the lack of age differences in processing times as typically older adults have longer reading times than young adults (Kliegl et al., 2004; Obler et al., 1991).

1.5.2. Resource allocation

The resource allocation approach was originally put forward in the 1980s (Graesser et al., 1980; Haberlandt et al., 1986; Haberlandt & Graesser, 1989; Just & Carpenter, 1980) but recently described in relation to language and aging by Stine-Morrow et al. (2006) who formulated the Self-Regulation Language Processing Model (see Figure 6).

Figure 6

Model of Self-Regulated Language Processing from Stine-Morrow et al. (2006)



The model reflects the architecture required to obtain an appropriate representation that includes, text, sentence and discourse levels and it is advocated that these processes work together to form a coherent discourse representation. Stine-Morrow et al. (2006) described the text level as representing orthographic decoding and lexical access. An example of sentence level processes is syntactic disambiguation. Finally, an example of discourse representation comprises of monitoring the protagonist within text.

The model has three assumptions relating to memory: 1) products in the model consist of negative feedback loops which must be activated in working memory to be executed, 2) long term memory can be accessed to obtain the appropriate goals and knowledge required for comprehension, and 3) when forming language representations, the goals and knowledge accessed in long term memory can be activated in working memory too. A further characteristic of the model is that the allocation of resources at the word and text level can become routine creating a 'habit of the mind'.

The reference value is affected by the attention required for the task. For instance, Stine-Morrow et al. (2006) suggest that there are a number of dependencies on the goals and values. For instance, in proof reading, the word and textbase reference values may be given more attention than a discourse level value (e.g., reading for pleasure). Furthermore, the cognitive, emotional, and social goals will also affect processing depending on the situation at hand. Social goals, such as, telling someone a story may result in discourse values being attended to. In contrast, cognitive and information goals, such as, reading to acquire knowledge may place emphasis on word and textbase values to acquire terminology and seek out details. Context and knowledge can affect processing at all levels.

In regard to aging, limited capacity in processing is thought to affect processing at all three levels of representation. Predominantly, age differences in reading occur due to older adults' inability to allocate resources to different components as efficiently as young adults, coupled with older adults' decline in executive function making it more difficult to engage the required resources. This fits with previous research showing that syntactically complex sentences and propositional encoding declines across the lifespan (Kemper et al., 2004). Another attribute of aging and reading in the Self-Regulated Language Processing Model is a reliance on the knowledge base, it is suggested that older adults are more likely to rely on social and emotional goals rather than the cognitive and information goals to compensate for age related decline.

Evidence supporting this Model of the Self-Regulated Language Processing comes from Stine-Morrow et al. (1996) who presented expository texts to young (M_{age} = 20) and older adults (M_{age} = 69) and used regression analyses to understand the allocation of resources young and older adults made to word, text and discourse features. Findings in the recall level amongst young and older adults were similar, however, differences in resource allocation were found. Specifically, older adults were more likely to utilise context compared to younger adults which was related to a higher recall (Miller & Stine-Morrow, 1998; Stine-Morrow et al., 1996). On the other hand, younger adults with high recall were focused on new concepts and word identification (Miller & Stine-Morrow, 1998; Stine-Morrow et al., 1996).

Additionally, Stine-Morrow et al. (2001) found that older adults allocated resources differently to their younger counterparts when reading. In the study, young and older adults read short passages of text twice with recall as the measure of comprehension. To control for any differences in recall, working memory was matched for young and older participants. Despite this, qualitative differences showed older adults were sensitive to task demands and were more likely to combine concepts within sentences. The authors argued that differences reflect general slowing during aging and older adults’ ability to compensate by allocating resources efficiently.

1.5.3. Risky reading strategy

Older readers of alphabetic languages, such as, English and German have been shown to change their reading capabilities in a variety of ways. To compensate for cognitive decline, on some occasions, older adults have been found to rely on context and their world knowledge employing a risky reading strategy where they make generally longer forward eye movements (McGowan et al., 2015; McGowan et al., 2014; Paterson et al., 2013a; Rayner et al., 2006, 2011, 2013) and skip words more frequently (McGowan et al., 2014, 2015; Paterson et al., 2013; Rayner, et al., 2006, 2011, 2013) but then look back to words more often when their inferences are incorrect (Kliegl et al., 2004; McGowan et al., 2014, 2015; Paterson et al., 2013; Rayner et al., 2006, 2011, 2013).

The term risky reading was first coined by O'Regan (1990). However, it was the influential paper by Rayner et al. (2006) who examined word frequency, word predictability and font difficulty in younger and older adults and found characteristics of risky reading, which instigated further exploration of this behaviour. In the study, Rayner et al. (2006) suggested that older adults compensated for slower lexical processing by relying on parafoveal information to guess upcoming words.

The risky reading strategy is particularly interesting because this reading pattern is not found in readers beginning to read (Reichle et al., 2013) and readers with reading difficulties, such as, dyslexia (Rayner 1985). Eye movements when reading for these particular groups comprise of more fixations on a word, more regressions and less skipping (Bellocchi et al., 2013). This suggests that risky reading characteristics are a *strategy* rather than a consequence of reading difficulty.

One explanation for older adults' risky eye movement behaviour when reading in English comes from research conducted on the perceptual span. The perceptual span is the available visual information during a fixation. In healthy young adult readers, the perceptual span tends to comprise of 3-4 characters to the left of the fixation and 14-15 characters on the right (McConkie & Rayner, 1975; Rayner et al., 1982; Rayner & McConkie, 1976). The benefits of the perceptual span is that depending on the words presented, readers can often begin processing upcoming letters of word(s) from the parafovea, parafoveal processing can speed up the processing of a word (Inhoff & Rayner, 1986; Schotter et al., 2012) and in some cases the word can be skipped altogether (Pollatsek et al., 2006; Reichle et al., 1998). Less research has been conducted into aging and the perceptual span. However, a series of studies by Rayner and colleagues suggest that older adults' perceptual span may be more symmetrical that young adult readers. Rayner et al. (2009) used the gaze-contingent paradigm to manipulate the amount of parafoveal processing young and older adults could engage in. The gaze contingent paradigm involves the letter outside a certain window being masked, this method provides useful information about the perceptual span and parafoveal processing. In Rayner at al.'s (2009) study, the x's masked letters of one (n+1) or two (n+2) words appearing after the target word (n). Results showed that young adults read faster when the window includes two words. Older adults, on the other hand, showed no preference, with significant disruption occurring in both conditions. Similarly, Rayner et al. (2010) used a moving window paradigm which consists of high-speed computing with extremely quick refresh rates that create an invisible boundary on a given target region where prior to crossing it, the boundary is masked and after it, it is not. Typically, readers are unaware of the boundary. Again, restricting upcoming information created less disruption for older adults than young adult readers. Rayner et al. (2009, 2010) concluded that older adult readers have a smaller, more symmetrical perceptual span, thus this group may engage in parafoveal processing less than their young counterparts. Due to older adults' inability to utilise parafoveal processing compared to young adults, it has been suggested that they are more likely to guess upcoming information (Rayner et al., 2006). However, not all the results from studies on parafoveal processing are consistent. Risse and Kliegl (2011) used the boundary paradigm to either show or mask the n+2 word. Young and older adults showed similar signs of disruption suggesting that older adults' perceptual span is preserved in later adulthood.

Furthermore, evidence by Whitford and Titone (2016) found no differences in age for young and older bilingual adults who read text which included spaces or where spaces were replaced by a dashes (-).

Additionally, risky reading has not been identified in all languages. Chinese older adult readers have instead been found to show a ‘cautious’ reading strategy where readers skip words less frequently (Wang et al., 2018; Zang et al., 2016). Although some research has questioned the characteristics of the risky reading strategy due to contrasting results (Choi et al., 2017), a meta-analysis of 22 studies found that younger adults made fewer regressions, skipped words less, had shorter gazes and read faster than older adult readers (Moreno et al., 2019).

Since Rayner et al.'s (2006) study, the characteristics of risky reading have been identified in a number of studies investigating lexical neighbours (Warrington et al., 2018), inter-word spacing (McGowan, White, Jordan, et al., 2014), implicit causality (Koornneef & Mulders, 2017), word masking (Rayner et al., 2011) and unspaced text (Rayner et al., 2013). Although numerous studies have identified a risky reading pattern in older adult readers, the potential consequences of risky reading for comprehension are not yet known as previous studies have investigated visual and basic word recognition factors affecting reading.

1.5.4. Good enough and shallow processing

Various strategies are described in the literature to try to explain differences in reading between young and older adults. What the strategies described above have in common is the role of older adults utilising experience and world knowledge acquired across the lifespan. There are two frameworks, good enough processing (Ferreira et al., 2002) and shallow processing (Sanford & Sturt, 2002) which advocate interpretations often result in a shallow, underspecified or partially processed interpretation. To avoid using additional resources, individuals rely on context and world knowledge when processing text and only conduct deeper processing of text if required. This framework and supporting evidence are described further in Chapter 2. In this thesis, I aim to investigate the age differences in shallow processing at the word, sentence and wider discourse level. Furthermore, an aim is to examine whether young and older adults engage in shallow processing online. Finally, I aim to examine the relationship between the two.

1.6. Cognitive factors, vocabulary and aging

Numerous language studies that investigate age differences have also examined how executive function may modulate language processing. Fluid intelligence, such as, working memory plays an integral role in the process of reading (Salthouse, 2012; Salthouse & Babcock, 1991). Furthermore, the speed of cognitive processing is likely to impact timely reading and comprehension as working memory has a limited capacity. In addition, skilled readers with a good vocabulary often read more quickly and possibly comprehend complex text more easily compared to individuals with lower vocabulary sizes (Ben-David et al., 2015). Working memory is integral to holding different representations and processing information within text. However, it is well known that various cognitive abilities declines during aging (see Salthouse, 2012 for a review). Despite a decline in fluid intelligence, crystallised intelligence has been found to be preserved and even increases across the lifespan (Gordon, Lowder, Hoedemaker, 2016). The next section explains in greater detail how these factors (working memory, processing speed and crystallised intelligence) influences older adults processing of text.

1.6.1. Working memory

Working memory is defined as the ability to store information whilst engaging in demanding tasks (Baddeley et al., 2009). Working memory is especially important in sentence comprehension as remembering the words that have come up before, being able to hold onto different representations and then integrating these representations with incoming information is integral for the development of a cohesive account of text (Just et al., 1996). The relationship between a lower working memory capacity and aging is consistent (Borella et al., 2008; Daneman & Merikle, 1996; Salthouse & Babcock, 1991). A consequence of aging is that older adults have lower working memory spans than younger adults (Salthouse & Babcock, 1991), this in turn has been found to affect this group's ability to retrieve information required to identify words (Daneman & Carpenter, 1980), as well as the capacity to process syntactically complex information (Christianson et al., 2006; Dede, 2014; Kemper et al., 2004; Van Dyke et al., 2014) affecting sentence comprehension (Daneman & Merikle, 1996).

When reading, individuals typically spend longer on the final word or phrase of a sentence or clause, it's been suggested that this phenomenon occurs due to integrative processing and is called clause- or sentence wrap-up (Just & Carpenter, 1980; Rayner et al., 2000; Warren et al., 2009). During language processing, time spent on the sentence wrap-up region has particularly been associated with working memory capacity (Miller & Stine-Morrow, 1998) as readers integrate the information within a sentence which often leads to a comprehensive representation of text. Payne and Stine-Morrow (2012) investigated age differences in semantic integration by using a boundary paradigm technique where a word ('N') was presented in a sentence-internal, clause-final, or a sentence wrap up region. The boundary paradigm technique enables the manipulation of the parafoveal preview benefit, enabling readers to see a parafoveal preview benefit or not. The parafoveal preview benefit is where information about an upcoming word ('N+1') is derived whilst the reader eye is focused on word N. When word N occurred in clause final and sentence wrap up regions, no age differences for early measures occurred meaning young and older adults had similar first fixation duration and single fixation duration times (Payne & Stine-Morrow, 2012). However, in later measures, young adults were found to have longer gaze duration and regression path reading times for the sentence wrap up region at N+1 than older adults. The authors suggested that older adults found it more effortful to conduct effective integration of concepts at the sentence wrap-up region (Payne & Stine-Morrow, 2012). A more recent study has found in addition to working memory, visual acuity and speed of processing to be factors relating to longer processing time at the sentence wrap-up region (Payne & Stine-Morrow, 2014). In contrast, higher verbal ability corresponded with faster reading times at the sentence wrap-up region (Payne & Stine-Morrow, 2014). These results show the complex interplay of various factors during online sentence processing.

In addition to clause- and sentence wrap-up effects, working memory has been found to influence processing of sentences containing temporary syntactic ambiguities. For instance, Christianson et al. (2006) presented young and older adults with garden path sentences, such as, '*While the man hunted the deer ran into the woods.'* The length of the ambiguous sentence e.g., '*While the man hunted the deer that was brown and graceful ran into the woods'* and ambiguity had no effect on age. However, verb type and plausibility did. Christianson et al. (2006) found in each experiment, working memory capacity modulated findings in age as older adults with greater working memory capacity were more likely to correctly answer questions (e.g., '*Was the man hunting the deer?*') related to understanding whether the reader had resolved the temporary syntactic ambiguity. Additional analysis found a strong positive correlation between working memory capacity and accuracy (Christianson et al., 2006). The authors concluded that older adults rely on heuristics to compensate for decline in working-memory capacity. The findings relating to working memory when reading temporary syntactic ambiguities has been mixed with some replication (Kemper et al., 2004) and other some studies finding no reliable working memory effects (Dede, 2014).

1.6.2. Processing speed

In addition to working memory, a decline in processing speed (Park et al., 1996; Salthouse, 1996) has been linked to healthy aging. Processing speed is simply the time taken to complete a task and has been found to mediate differences found in working memory (Park et al., 1996) as the speed that information is computed affects the time taken to comprehend text (Borella et al., 2011). Salthouse (1996) suggests that older adults' ability to process information affects their ability to store incoming text and remember previous information, this in turn affects their ability to process sentences, this has been termed the processing speed theory of adult age differences by Salthouse (1996).

Waters and Caplan (2005) conducted a battery of processing speed tasks as well as different sentence processing tasks on young and older adults. The sentence processing tasks included an online listening task, and end-of-sentence plausibility judgement task and discourse processing using the Nelson Denny comprehension test. Results showed that older adults had slower processing speed on the plausibility judgement task. Furthermore, processing speed only moderated processing of cleft-object sentences (e.g., '*It was the movie that terrified the child because it showed a monster.*') and not sentences with less complex structures. It-cleft sentences pose difficulty during language processing because the thematic roles of a sentence are violated as the sentence structure does not adhere to the noun-verb-noun structure. In contrast, many reading studies have found older adults to have generally longer reading times, this often allows readers time to integrate information needed to correctly comprehend text (Kemper et al., 2008; Rayner, Reichle, et al., 2006). Furthermore, eye-tracking studies have found older adults to have longer reading times, specifically at the wrap up region and age differences in processing speed (and working memory) accounted for these extreme times (Payne & Stine-Morrow, 2014; Staub & Benatar, 2013).

1.6.3. Vocabulary

Despite a decline in working memory, processing speed and vision across the lifespan, crystallised intelligence is one area that has been found to be preserved during aging, and in some circumstances, older adults have been found to show an advantage compared to younger adults (Brysbaert et al., 2016). Crystallised intelligence involves the acquisition of knowledge and experience that expands across a lifetime of reading (Salthouse, 2009) and this may preserve sentence processing until later adulthood (Paterson et al., 2020; Salthouse, 2009). For instance, a longitudinal study testing 2000 young (18-35 years) and older (60-94 years old) adults over 16 years on vocabulary performance found older adults always outperformed young adults when education was controlled (Keuleers et al., 2015). Further research has found vocabulary size to increase and age was the biggest predictor of vocabulary size accounting for 17% of variance (Ben-David et al., 2015). It has been suggested that words are often used in different contexts which reinforces the acquisition of that word in one's vocabulary (Burke, MacKay & James, 2000). Furthermore, learning words via experience and knowledge amplifies the memory trace created (Burke et al., 2000). Thus, older adults' extended vocabulary size may enable this group to mitigate against deterioration in fluid intelligence.

1.6.4. Summary

In the experiments reported in this thesis, working memory and vocabulary were assessed. The digit-span forward, digit-span backward and the digit-span sequence tasks were used to assess working memory as this test is one of the most widely used tests for working memory (Wechsler, 2008). Composite working memory scores of the Weschler Adult Intelligence Scale (WAIS-IV) were calculated by adding scores from the individual tests. Processing speed was measured using the Weschler Adult Intelligence Scale (WAIS-IV).

Vocabulary was measured using the Weschler Adult Intelligence Scale (WAIS-IV), specifically the vocabulary sub-test was administered to participants. During the test, they were required to define up to 30 words, correctly doing so for each word produced a maximum score of 50 points. The aim was to run exploratory analyses adding these variables in to the analyses to better understand the role of working memory, processing speed and vocabulary in language processing.

In addition to working memory measures, two cognitive assessments were administered to all older adult participants. The Montreal Cognitive Assessment (MoCA) is a 10-minute assessment that aims to assess signs of mild cognitive impairment by assessing multiple cognitive domains, including: attention, concentration, executive functions, memory, language, visuospatial skills, abstraction, calculation and orientation (Nasreddine et al., 2005). The highest score on the MoCA is 30 and an inclusion criterion of ≥ 26 was applied, which is in accordance with the guidelines. The Mini Mental State Examination is the most common cognitive impairment screening procedure (Folstein et al., 1975). This assessment takes around 5-10 minutes and examines orientation, memory, attention, and verbal and written commands.

1.7. Conclusion

This chapter outlined the differences in aging during reading. Specifically, key age differences that occur during reading at the word-, sentence- and wider discourse-level were described. This chapter also described the strategies that older adults may use to compensate for difficulty during reading. What these strategies have in common is a reliance on world knowledge and experience. Two frameworks, shallow (Sanford & Sturt, 2002) and good enough (Ferreira et al., 2002) processing advocate that readers utilise *fast and frugal* heuristics for language comprehension. As a result, processing of text often results in a shallow, underspecified or partially processed representation. The shallow and good enough processing frameworks and supporting evidence are described in the next chapter.

The shallow and good enough processing frameworks are used to assess the role of world knowledge and experience in young and older adults during reading with manipulations at the word, sentence and wider discourse level. One of the aims of this thesis is to better understand the role of context and the potential consequences for language comprehension for older adults when compared to their younger counterparts, whilst accounting for any visual and cognitive moderating factors.

Chapter 2: Shallow processing

Shallow processing differs from other reading models (Ferreira & Clifton, 1986; Frazier & Rayner, 1982; MacDonald et al., 1994; Tanenhaus et al., 1995) in that this theoretical framework advocates that a full syntactic and semantic analysis may not always be accomplished, instead readers often rely on context and their knowledge of the world (Ferreira et al., 2002; Sanford & Graesser, 2006; Sanford & Sturt, 2002). Shallow processing often leads to an interpretation which is partial or incomplete and thus remains underspecified (Sanford & Graesser, 2006). Sanford and Sturt (2002) claim that shallow processing occurs when a full analysis is not conducted, which in many situations, they suggest, is not necessary. For example, if a teacher is asked by a mother about her child's performance 'Did my kid manage to climb up a tree?', and the teacher responds 'Right now, every kid is up a tree!', the mother will by inference know that if every kid is up a tree then her child is up a tree (Sanford & Graesser, 2006). The exact meaning as to whether every kid is up the same tree or a different tree will not be important and so in this instance, deep processing is not necessary.

Good enough processing is a related framework that similarly suggests processing is often shallow, which leads to a good enough representation as opposed to a detailed representation (Ferreira et al., 2002). More recently, Karimi and Ferreira (2016) reviewed research relating to good enough processing and extended the framework to suggest that algorithmic processing may occur if the task requires it and this can be conducted in parallel to good enough processing or algorithmic processing may be delayed depending on task requirements. Good enough processing is believed to be underpinned by a *fast and frugal* heuristic which often results in a faster and good enough result.

The idea that heuristics override computational processing originates from the broader behavioural economics literature. The school of thought for decades in judgment and decision making was that humans are rational agents and so complete algorithmic processing takes place in order to resolve judgements and decisions in one's environment (Goldstein & Gigerenzer, 2002; Mousavi & Gigerenzer, 2014). However, this view has been contested, most notably by Nobel prize winners, Kahneman and Tversky who showed that instead of algorithmic processing, individuals often rely on heuristics (Tversky & Kahneman, 1989). Heuristics are quick, simple and reliable tools, often referred to as *rules of thumb* or *shortcuts* which are available and resolve complex problems quickly and efficiently (Mousavi & Gigerenzer, 2014). The *fast and frugal* heuristic that is suggested to underpin good enough processing is an umbrella term for a range of heuristics that operate in limited time, information and knowledge in place of complex algorithms often resulting in processing that is superficial (Mousavi & Gigerenzer, 2014). Throughout this thesis, the terms shallow and good enough processing are used synonymously. The remainder of the chapter focuses on key evidence of shallow processing at the word, sentence and wider discourse level.

2.1. Shallow processing and anomaly detection (word level)

Some of the most compelling evidence of shallow processing has been identified at the word level, using semantic anomalies. Whilst some anomalies like '*He drinks his tea with dog'* are instantly noticeable, other semantic anomalies often go unnoticed. For instance, when individuals are asked '*How many of each sort of animal did Moses put on the Ark?',* the common answer is *'two'.* Despite knowledge that it was Noah on the ark and not Moses, the semantic anomaly '*Moses*' is not identified as anomalous (Erickson & Mattson, 1981). Erickson and Mattson (1981) coined this type of semantic illusion as the *Moses illusion*. Semantic illusions occur when a sentence is syntactically correct, however a portion of text within the sentence is incorrect rendering the interpretation as entirely different. This illusion is considered an example of 'shallow processing' because it highlights that when a target word fits with contextual information, as well as the world knowledge of the reader, anomalies often go undetected (Bohan & Sanford, 2008). For example, Barton and Sanford (1993) presented participants with passages such as (1) and found readers only detected that the word 'survivors' was anomalous 33% of the time, compared to higher detection rates for 'injured' (66%). They hypothesised that this result occurred because of a core assumption of the meaning of the word 'survivors' (1a), in that it refers to an individual who has managed to live despite others dying from an event. In contrast, the definition of 'injured' (1b) relates to someone being harmed but there is no association with death. However, it is not only the word that affects detection, 'survivors' also fits with the global context of an air crash. When participants were asked '*When a bicycle crash occurs, where should the survivors be buried?*' detection rates rose to 80% (Experiment 3, Barton & Sanford, 1993).

1. There was a tourist flight travelling from Vienna to Barcelona. On the last leg of the journey, it developed engine trouble. Over the Pyrenees, the pilot started to lose control. The plane eventually crashed right on the border. Wreckage was equally strewn in France and Spain. The authorities were trying to decide where to bury the survivors/injured.

a) After an air crash, where should the **survivors** be buried?

b) After an air crash, where should the **injured** be buried?

2.1.1. Semantic relatedness

The ability to detect semantic anomalies is not only dependent on the context of the sentence, but how related the anomalous and non-anomalous versions are. For instance, in the traditional example, known as the Moses illusion '*How many of each sort of animal did*

Moses put on the ark?', 'Moses' is highly related to 'Noah' as Moses and Noah are both significant biblical figures, thus the detection rate for 'Moses' was low (49%). A similar result occurred for other biblical figures, such as, Abraham with detection at 44%. When Noah was replaced with a less semantically related name but high phonological similarity, detection rates were high. 'Nixon' a former president who has the same number of syllables, the same initial sound and same syllable stress but low semantic similarity replaced Noah and detection rates increased to 100%. Erickson and Mattson (1981) attributed this high detection rate to a lack of semantic relatedness. However, this study cannot rule out the influence of global context as the low semantic related condition (e.g. *Nixon*) did not fit with the global context (biblical subject) of the sentence.

More recently, Cook et al. (2016) used eye-tracking to investigate semantic relatedness by presenting true/false statements, for example, '*Dogs usually don't like the sound of thunder during a storm*' to participants, which contained anomalies that were likely ('*lightning*') or unlikely ('*sunshine*') to be related to the correct non-anomalous information ('*thunder*'). The experiment also compared processing differences between detected and undetected semantic anomalies. Longer first-pass reading times were found for unlikely anomalies '*sunshine*'. Furthermore, unlikely ('*sunshine*') and likely ('*lightning*') anomalies incurred longer second-pass reading times compared to non-anomalous sentences ('*thunder*').

Cook et al. (2014; 2016) postulated an alternative model to the shallow processing framework. The RI-Val model includes three stages; the *Resonance* stage is the process of incoming information being encoded with information stored in the long-term memory; information processed in the resonance stage is then integrated with information in the working memory - *Integration stage*; then at *Validation* – the interpretations derived are validated using a low-level partial matching procedure. Processing is complete when the processing is considered 'good enough' via a *coherence threshold.*

This study is particularly interesting as it complements offline tasks showing that a lack of semantic relatedness results in higher detection rates as more unrelated anomalies were identified early in processing in Cook et al.'s study. The RI-Val model suggests all anomalies are processed in the same manner as detected anomalies, however Resonance, Integration, and Validation stages occur late for undetected anomalies, resulting in delayed processing difficulty.

2.1.2. Focus

Focus is another factor that has been found to affect depth of processing of semantic anomalies. Bredart and Modolo (1988) used sentences with a narrow focus (e.g., 2) containing it-clefts and these types of sentences resulted in higher detection rates (91%) than sentences with a broad focus, such as the original sentence (66%) (3) used in Erickson and Mattson's (1981) study.

2) It was Moses who put two of each kind of animal on the Ark.
3) Moses put two of each kind of animal on the Ark.

It-clefts resulted in a greater depth of processing of the target word '*Moses*' as the sentence was effectively answering the question '*Who was it that put two of each kind of animal on the Ark?'*. Typography can also increase focus, for instance, including the anomalous word in uppercase letters and underlined ('*<u>MOSES</u> decided to take two animals of each kind on the Ark')* increased the anomaly detection rate (86.5%) compared to underlining and capitalising other information ('*Moses decided to take <u>TWO</u> animals of each kind on the Ark*') (68.3%) (Bredart & Docquier, 1989).

Additonally, Sturt et al. (2004) conducted two experiments using a text-change detection paradigm to examine the role of syntax and focus on depth of semantic processing. In Experiment 1, Sturt et al. (2004) presented an initial display of the item for 8 seconds to half of the participants, whilst the other half of participants read the item at their own pace. After the initial display, a grey screen appeared for a duration of 500 msec, which followed a second display of the item for 10 seconds. Participants were tasked with detecting whether there was a change at the critical noun in (4a and 4b) from ‘cider’ to a semantically related meaning ‘beer’ or a semantically unrelated meaning ‘music’ between the initial and subsequent display of text. When the meaning was related, a change was detected more often when the critical word was in focus (4a) compared to when it was not (4b). However there was no difference in detection when it was an unrelated meaning.

4) Everyone had a good time at the pub. A group of friends had met up there for a stag night.

4a. What Jamie really liked was the cider $_{\text{(focus word, target word)}}$, apparently.

4b. It was Jamie who really liked the cider $_{\text{(unfocused word, target word)}}$, apparently.

In Experiment 2, sentences were presented to participants either in fous (5) or not in focus (6). Sentences like (5) were in focus because the the embedded question (‘*which man was got in to trouble?*’) implicitly prepares the readers to attend to (‘*with the hat*’*)* in the subsequent sentence. On the other hand, the unfocused context (6) has no emphasis on any particular individuals, providing a broad context. Semantic distance was also manipulated so that in the text-change detection task, ‘*hat*’ would change to a semantically related catgeory, such as, ‘*cap*’ or to an unrelated category, like ‘*dog*’.

5) Everybody was wondering which man got into trouble. (focused context) In fact, the man with the hat (target word) was arrested.

6) Everybody was wondering what was going on that night. (unfocused context) In fact, the man with the hat (target word) was arrested.

It was expected that participants would notice changes to the target word in the focused condition when a semanitcally related target appeared than the unfocused condition and this is what was found. Together, these experiments provide evidence for shallow processing, indicating that inducing deep processing through focus via syntax and context can reduce shallow processing. The authors further suggested that the findings can be explained by the granularity hypothesis (Hobbs, 1985), which advocates that there are multiple gradings of semantic meaning. If a reader derives a coarse granularity, then in the change-detection task ‘hat’ and ‘cap’ maybe underspecified as head wear (Bohan & Filik, 2018; Hobbs, 1985; Sturt et al., 2004). Sturt et al. (2004) explained that linguistic focus can result in readers conducting a more sophisticated represenation. Note that evidence against focus inducing greater graularity has recently been found (Bohan & Filik, 2018).

2.1.3. Processing load

Processing load has also been found to affect detection of semantic anomalies. Bohan (2007) investigate the influence of processing load. To do this, Bohan (2007) presented items where the text prior to the target word (*hostages*) contained a lot of additional information - high load condition (7) or little information - low load condition (8).

7) Question: Under these circumstances, what difficulties would the officials at John F. Kennedy Airport, who must ensure the safety of their passengers, be facing when they must negotiate the demands of the **hostages** and possible further threats to airport security? (High load)

8) Question: Under these circumstances, what difficulties would the officials at John F. Kennedy Airport, who must negotiate the demands of the **hostages**, be facing when they must ensure passenger safety and possible further threats to airport security? (Low load)

In part one, participants read a booklet containing semantic anomalies and were asked to read for comprehension but to indicate if any anomalies occurred (as anomalies had been spotted previously and an example was given). In part two, participants reread the booklet and again were tasked with identifying anomalies. In part three, only items with anomalies were provided and the anomaly was presented in bold with an explanation as to why the item was anomalous. Participants indicated whether the anomaly was spotted in part one, two or three. As expected, detection rates were lower for the high load (31%) (part one: 10%, part two: 21%) versus the low load (40%) condition (part one: 10%, part two: 30%).

Similarly to Sturt et al. (2004), Sanford et al. (2005) used a text-change paradigm task to investigate the effect of processing load on sentence comprehension. Syntactic complexity (Experiment 1), higher referential load (Experiment 2) and referential complexity (Experiment 3) were found to result in lower detection rates. However, Experiment 4 found that a high referential load does not impede comprehension as differences in between low (97%) and high (93%) referential loads were statistically non-significant. Interestingly, results showed that processing load effects were localised to the embedded verb (Experiment 5). Although referential load and focus affected change detection, results indicated that the effect processing load had on change detection differed to the effect of focus. Focus highlights a certain area of text, thus affecting the depth of semantic processing. On the other hand, sentences that contained a high load resulted in greater difficulty at the embedded verb, and so the results instead suggest that accessing the memory trace of the word may be the reason that load affects detection rates.

2.1.4. Unconscious processing of semantic anomalies

Bohan and Sanford (2008) conducted an eye-tracking study and found no differences in eye movements for undetected anomalies and non-anomalous sentences (full description can be found in Chapter 4). This study was followed up with an ERP experiment where anomalies that were a poor fit (9) or a good fit (10) to the context were presented (Sanford et al., 2011). Sentences were presented that did not fit with the context and were deemed a 'poor fit' (9) condition or sentences did fit with the context of the discourse and were categorised as a 'good fit' (10) condition. For instance, the first sentence provided context and in (9) '*letters*' does not fit with the global context of a record shop. Furthermore, in (9) '*letters*' does not fit with the local context as letters are not sold. In (10) the anomalous word, 'sentence' fits with the global context of a child abuse court case. Furthermore, in (10) 'sentence' provides a local fit for the target word 'victim'.

9) Leon was the manager of a struggling record shop. Yesterday, the owner told him that he would have to think of new ways to sell more {letters (anomalous word) / records (non-anomalous word)} if he wanted to keep his job. ('Poor fit' condition)

10) Child abuse cases are being reported much more frequently these days. In a recent trial, a 10-year {sentence (anomalous word) / care order (non-anomalous word)} was given to the **victim (target word)**, but this was subsequently appealed. ('Good fit' condition)

The N400 ERP component is associated with semantic inconsistencies and was a component of interest within the study. Results from this study showed that N400 effects occurred when easy-to-detect semantic anomalies that had a poor fit to the context (9) were presented. However, an N400 effect was not elicited for hard-to-detect semantic anomalies, which were a good fit to the context (10). Interestingly a larger late posterior positivity (LPP), specifically in the P600 range was found for both types of anomalies compared to non-anomalous sentences and no differences in ERPs were found for undetected anomalies and non-anomalous sentences. The authors suggested that the greater LPP activity in the P600 range reflected a more general violation between syntactic interpretation and semantic violation. This study was taken as evidence of shallow processing as no significant differences between undetected and non-anomalous sentences were identified (Bohan & Sanford, 2008; Sanford et al., 2011). Thus, the ERP findings cannot be due to reduced awareness hypothesis as the ERP activity for good-fit undetected anomalies was similar non-anomalous sentences suggesting that the anomalies were not registered.

As mentioned above, Cook et al. (2016) used eye-tracking to investigate processing of semantic anomalies using true/false statements, for example, '*Dogs usually don't like the sound of thunder during a storm*'. The experiment also compared processing differences between detected and undetected semantic anomalies. Unlikely '*sunshine*' and likely '*lightning*' anomalies incurred longer second-pass reading times compared to non-anomalous sentences '*thunder*'. Thus, this study found differences in online processing between detected and undetected semantic anomalies when compared to the non-anomalous sentences suggesting processing of undetected semantic anomalies may illustrate unconscious registration (Cook et al., 2016; Reder & Kusbit, 1991).

Further evidence for unconscious processing of semantic anomalies comes from Hannon (2015) who used the moving-window paradigm, which tracks the eye movements of participants but controls the amount of information a reader can process. In this experiment, the reader could only process the word that they were fixated on. Sentences which included anomalies were presented (e.g., '*What superhero is associated with bats, Robin, the Penguin, Metropolis, Catwoman, the Riddler, the Joker, and Mr. Freeze?*'). In the previous example, Metropolis is not a superhero, however this anomaly often goes undetected by readers. The study found longer total reading times for detected compared to undetected anomalies at the critical region. Furthermore, longer reading times were found for anomalous sentences that went undetected compared to non-anomalous sentences on the post-critical regions. The authors concluded that the detection of anomalies influences the language processer to immediately resolve the incongruent information. On the other hand, when anomalies are not consciously detected, the language processer still aims to resolve the contradictory information, however, this is delayed. The shallow processing theory was not discounted, Hannon (2015) suggested three possible explanations for non-detection: firstly, shallow processing may take place where the full meaning of the word is not accessed and/or this is not integrated with the wider discourse. In accordance with the two cognitive mechanism theory (Daneman & Hannon, 2001) there may be issues with semantic integration of undetected anomalies and knowledge processes which in turn affects comprehension. Another explanation is that the processing of the meaning of the anomalous word is completed but integration with the rest of the text is not consciously processed, this hypothesis is referred to as the *reduced awareness hypothesis* (Sanford et al., 2011).

In conclusion, whilst there is some compelling evidence for shallow processing, more recent evidence indicates unconscious processing of semantic anomalies may be conducted by individuals, this is an area that is explored further in Chapter 4.

2.1.5. Individual differences

A range of different factors have been found to affect processing of semantic anomalies, however, an area that has received less attention is how individual differences may play a role in the detection rate of semantic anomalies. Hannon and Daneman (2001) presented undergraduates with anomalous questions (11 and 12) that were higly semantically related (e.g. *holiday*) (11) or less semantically related (*occassion*) (12) to the answer (*Halloween*).

11) On what holiday do children go door-to-door, dressed in costume, *giving out* / *buying* candy?

12) On what occasion do children go door-to-door *giving out* / *buying* candy?

Furthermore, working memory was examined using the Daneman and Carpenter reading span test and a knowledge test that assessed participants' knolwedge of the item (e.g., Halloween). Additionally, a knowledge access test was conducted, this test measured: a participant's capability in accessing knolwedge, integrating this knolwedge with incoming text, the ability to make inferences and finally, the recall of text from long-term memory. Similarly to previous studies examining semantic anomalies, the more related the anomaly was to the non-anomalous word and the global context, the more likely semantic anomalies were to go undetected. Furthermore, regression analyses indicated that participants' knowledge access accounted for a higher variance of detection (23%) than working memory capacity (19%). Both variables were additive and the authors concluded that both knowledge access and working memory are individual differences that contribute to the ability to detect semantic anomalies.

In a further study, Hannon and Daneman (2004) split participants in to readers who were high (above the 50th percentile) or low skilled (up to the 50th percentile) based on results from the Nelsen-Denny test, a reading comprehension assessment. Results showed that low-skilled readers had lower detection rates for sentences containg semantic anomalies (28%) compared to high skilled readers (60%). Accordingly, low-skilled readers engaged in shallow processing more than high skilled readers resulting in this group being less able to detect semanic anomalies. However, this study did not examine knowledge and so whether this was controlled for participants is unknown.

Further evidence of individual differences affecting processing of anomalies comes from literature on individuals with Autism-Spectrum disorder (ASD). Au-Yeung et al. (2018) explored the role of context at the paragraph (13) and sentence (14) level in young adults with ASD and typically developing adults using eye-tracking methodology.

13) After three years of hard work on his degree in Oriental Studies, Scott finally graduated from university. After his graduation, he received a job offer to work in Japan for a year. He was really excited about this and he was planning to take the opportunity to travel around Tokyo, one of the most vibrant cities in East Asia. However, Scott was worried that his inability to speak **Japanese/Chinese** would stop him from communicating with people. He really wanted to be able to make new friends out there. (Global context)

14) There was a tourist flight travelling from Vienna to Barcelona. On the last leg of the journey, it developed engine trouble. Over the Pyrenees, the pilot started to lose control. The plane eventually crashed right on the border. Wreckage was equally strewn in France and Spain. The authorities were trying to decide where to bury the **dead/survivors** from the plane crash. The families of the passengers were devastated about their losses. (Local context)

The results from this study showed that in the paragraph condition (13), participants were required to process the global context of the whole paragraph to complete successful anomaly detection. In contrast, at the sentence level (14), violations of a thematic role detected within a sentence were required. Participants with ASD identified local-context anomalies earlier than typically developing individuals and showed evidence of immediately checking the anomaly. In contrast, in the global context anomaly condition, typically developed participants identified anomalies earlier and spent more time looking back at the anomalous region than ASD participants. Thus, the results were reversed by condition. The authors suggested that individuals with ASD experience greater difficulty connecting incoming information with global context, compared to typically developing individuals (Au-Yeung et al., 2018).

In sum, many studies using a variety of methodologies have identified that semantic anomalies often go undetected, despite the reader's knowledge. Some of the main findings identified include that semantic relatedness and the global and local context of the scenario can affect a reader's ability to detect semantic anomalies. Additionally, focus and load can affect the depth of processing of anomalies. Another crucial aspect to explore is whether undetected anomalies are registered even if they are not reported. It is suggested that semantic anomalies go undetected because depth of processing is shallow and results from studies investigating focus have found that focus improves detection of semantic anomalies. This is important to explore because if undetected semantic anomalies are implicitly processed it raises the question as to whether semantic anomalies go undetected due to shallow processing. This is explored further in Chapter 4.

Previous studies have identified individual differences in the processing of semantic anomalies. Specifically, the ability to access knowledge and cognitive factors, such as working memory contribute to the ability to identify semantic anomalies. Furthermore, differences between typically developing readers and readers who have ASD occur possibly due to an inability to generate a complete and detailed discourse description during reading. Although changes in cognition occur across the lifespan, crystalised intelligence has been found to remain stable and even expand (Gordon, Lowder & Hoedemaker, 2016). More world knowledge and experience are acquired across the lifespan, therefore, older adults may utilise this during reading. If older adults do rely on context, this group may be more likely to engage in shallow processing compared to their younger counterparts. One study to date has investigated differences between young and older adults in processing semantic anomalies, this is described in Chapter 4.

2.2. Shallow processing and temporary syntactic ambiguities (sentence level)

The occurrence of shallow processing at the sentence level has been widely established using sentences containing temporary syntactic ambiguities. Temporarily ambiguous sentences are often referred to as *garden path* sentences because the first part of the sentence influences the reader to form an incorrect interpretation, however, if the language comprehension system reprocesses the sentence, the correct interpretation can be adopted. For example, readers often initially form an interpretation of sentences like (15) as *Anna dressing the baby.* However, when encountering the verb *played*, a reanalysis of *the baby* should occur. Instead of forming an interpretation of *the baby* as the object of *while Anna dressed*, *the baby* should be processed as the subject, forming an interpretation that the baby played in the crib whilst Anna dressed herself. Often readers do not engage in deep processing, instead opting for a shallow or a good enough interpretation, this initial incorrect representation (*Anna dressing the baby*) can then linger. As a result, when asked questions (e.g., *Did Anna dress the baby?*), individuals will incorrectly answer ‘*yes*’. When a comma is positioned after the noun phrase (16), it becomes clear that Anna is dressing herself, therefore the inclusion of a comma prevents ambiguity in these types of sentences (Christianson et al., 2001; 2006).

15) While Anna dressed the baby played in the crib. (ambiguous sentence)

16) While Anna dressed, the baby played in the crib. (unambiguous sentence)

Did Anna dress the baby?

In a series of influential experiments, Christianson et al. (2001; 2006) explored the factors that may lead young adults (2001) to fail to resolve temporary syntactic ambiguities. One factor that can create difficulty in processing of temporary syntactic ambiguities is the distance between the ambiguous region and syntactic disambiguation phrase (Ferreira & Henderson, 1991). This was manipulated by presenting short sentences containing temporarily syntactic ambiguities (a and c) and long sentences with temporary syntactic ambiguities (b and d) (Christianson et al., 2001). Another interesting factor explored was the pragmatic plausibility of temporarily ambiguous sentences (Christianson et al., 2001). For example, in Table 1, based on context and world knowledge, it is more plausible that a deer is being hunted in the park than a zoo. Results found the shorter the ambiguous region (a and c) and the more implausible the region (c and d) resulted in more participants correctly answering '*No*' to garden path questions (Experiments 1a and b, Christianson et al., 2001). These results suggest that readers engage in good enough processing, placing a greater reliance on context during the processing of temporary syntactic ambiguities (Pickering & Traxler, 1998). Consequently, confidence ratings were measured and ratings for incorrect 'yes' answers were similar to correct 'no' answers to comprehension questions (see Table 1). Furthermore, directing the question at the subordinate clause (*Did Bill hunt the deer?*) compared to the main clause (*Did the deer run into the woods?*) resulted in lower accuracy rates (Experiment 2) (2001).

Table 1

Example temporarily ambiguous sentences in Christianson et al. (2001)

Number	Condition	Example sentence
a.	Plausible, short ambiguous	While Bill hunted the deer ran into the woods.
b.	Plausible, long ambiguous	While Bill hunted the deer that was brown and graceful ran into the woods.
c.	Implausible, short ambiguous	While Bill hunted the deer paced in the zoo.
d.	Implausible, long ambiguous	While Bill hunted the deer that was brown and graceful paced in the zoo.
e.	Plausible short ambiguous	While Bill hunted the pheasant the deer ran into the woods.
f.	Implausible long ambiguous	While Bill hunted the pheasant the deer that was brown and graceful ran into the woods.

Question Number	Clause	Question
1	Subordinate clause	Did Bill hunt the deer?
2	Main clause	Did the deer run into the woods?

Christianson et al. (2001) concluded that the misinterpretation of temporarily ambiguous sentences is due to a lingering misinterpretation that is often not successfully reanalysed with partial processing taking place meaning that the initial interpretation of '*Bill' hunting the deer* is never fully rejected. Christianson et al. suggested two further underlying explanations: a) reanalysis is never fully completed or b) both interpretations are blended. In relation to the former explanation experiments pertaining to online processing of temporary syntactic ambiguities have aimed to examine this. For instance, Slattery et al. (2013) investigated whether misinterpretations of sentences containing temporary syntactic ambiguity is due to incomplete reanalysis or the failure to build a correct structure. In Experiment 1, gender mismatch effects were exploited to examine the structure built when encountering sentences containing temporary syntactic ambiguities. Gender mismatch effects describe the established costs that occur because readers immediately search for the antecedent of a reflexive pronoun if the gender does not match (e.g., 18) (see Slattery et al., for a detailed explanation). Sentences containing temporary syntactic ambiguities (17 and 18) were presented where initially readers may interpret the sentence as '*The bank manager called David's father/mother*'. Furthermore, the reflexive '*himself/herself*' either matched or mismatched the definitional gender '*mother/father*'. Unambiguous controls included the same words as the ambiguous condition but a comma was included after '*telephoned*'.

17) After the bank manager telephoned David's father grew (disambiguation region) worried and gave himself (reflexive) approximately five days to reply. (Ambiguous match)

18) After the bank manager telephoned David's mother grew (disambiguation region) worried and gave himself (reflexive) approximately five days to reply. (Ambiguous mismatch)

The results showed a clear garden path effect at the disambiguation region '*grew*' with longer first-pass, go past and total reading times for ambiguous sentences than unambiguous controls. However, at the reflexive 'himself/herself', no garden path effect was found. This finding suggested that a hierarchical syntactic structure had been derived by the time the reflexive '*himself/herself*' was encountered. Furthermore, results showed at the reflexive region 'himself / herself' longer first-pass, go past and total reading times for the gender mismatch (18) compared to gender match (17). As found previously, gender mismatch effects a cost occurred. In Experiment 2, eye movements during reading were again utilised to explore for the linger misinterpretation. Two sentences were presented to determine if there was interference from the syntactic ambiguity in the first sentence on the second sentence. Results showed that a significant cost in processing the second sentence was evident. Slattery et al. advocated that the study was evidence of 'good enough' processing as lingering misinterpretations were found. Furthermore, the authors suggested that incomplete processing occurs due to an inability to inhibit the initial garden-path interpretation, which results in two representations, the initial garden path structure and the representation formed during reanalysis.

One study has attempted to examine whether sentences containing temporary syntactic ambiguities result in blended interpretations in young and older adults (Malyutina & den Ouden, 2016). Young and older adults listened to sentences containing temporary syntactic ambiguities similar to those used in Christianson et al.'s (2001) study (see Table 1) and a multiple-choice sentence picture matching task was used to assess participants' representations. Black and white line drawings were used to represent the correct interpretations (Bill hunting a bird and a deer running into the woods), the garden-path representation (Bill hunting the deer and a different animal running into the woods) and then a blended interpretation (Bill hunting the deer which is running into the woods). Results indicated that younger adults were more likely to opt for blended choices and older adults were more likely to choose garden-path representations. Malyutina and den Ouden (2016) suggested that older adults may be less likely to incorporate incoming information due to a low working memory capacity or slower processing speed.

Literature has consistently shown that both young and older adults engage in shallow processing when processing temporary syntactic ambiguities, resulting in a failure to reanalyse the misinterpretation correctly. Studies have examined the underlying process of parsing sentences containing temporary syntactic ambiguities, whilst it is clear that a representation if formed, it is less clear whether representations are blended or whether there are two separate representations formed. The type of verb used (e.g., optionally transitive and reflexive absolute transitive), the more plausible the context of the sentence is and question type can affect depth of processing temporary syntactic ambiguities. Numerous studies have suggested cognitive factors: working memory, processing speed and inhibitory control may moderate the ability to process syntactic temporary ambiguities correctly.

Shallow processing is a framework which advocates that readers utilise world knowledge and experience and rely on context rather than completing a full syntactic analysis on temporarily ambiguous sentences. As more knowledge and experience is accumulated across the lifespan, older adult readers may be more likely to engage in shallow processing when processing temporarily ambiguous sentences. Christianson et al. (2006) expanded on their work with younger adults (Christianson et al., 2001), using a self-paced reading task and investigated sentence structure (garden path versus non-garden path), plausibility, verb type (RAT versus OPT verb), length of the sentences and working memory capacity of young and older adult readers. The distance between the ambiguous information and disambiguation region had no effect on temporary syntactic ambiguity disambiguation. Furthermore, the plausibility of the sentence and the sentence structure did not affect syntactic ambiguity disambiguation. Older adults were more likely to be garden pathed when sentences contained an OPT verb compared to young adults, Christianson et al. (2006) concluded that this relates to semantic inferences older adults hold and suggested that good enough processing is one of many heuristics that older adults possess. Furthermore, working memory capacity was critical as participants with low working memory capacity were more likely to be garden pathed. This result is unsurprising as sentences containing temporary syntactic ambiguities require the reader to create a syntactic structure and hold onto a number of thematic roles. Aging and processing of sentences containing temporary syntactic ambiguities is explored further in Chapter 5.

2.3. Shallow processing and pragmatic normalisation (sentence level)

So far, this chapter has focused on syntactic ambiguity and anomalous sentences, however, there are instances when individuals engage in shallow processing with unambiguous sentences. Fillenbaum, (1971; 1974) first identified this and conceptualised it as ‘pragmatic normalisation’. In his studies, disjunctive sentences were presented which included conditional threats e.g. ‘*Clean up the mess or I won’t report you’*. Often it was found that readers normalised these sentences ‘*If you don’t clean up the mess I’ll report you*’ instead of preserving the original, albeit unusual meanings ‘*If you clean up the mess I’ll report you*’. Furthermore, when questioned by the experimenter, participants were often unaware that they had changed the meaning of these types of sentences. This has now been coined the ‘depth charge’ illusion (Wason & Reich, 1979).

The focus of pragmatic normalisation in this thesis is in relation to differences in canonical and non-canonical sentence structures. Shallow processing has been found to occur even in sentences void of any ambiguity or anomaly. For instance, in Ferreira's (2003) study, participants listened to sentences presented in an active (19 and 20) or passive (21 and 22) structure and answered who the do-er was in each sentence e.g., '*Who did the biting?*'. In addition to sentence structure, plausibility was manipulated. Comprehension rates were 96% for active plausible sentences (19), 95% for active implausible sentences (18), 92% for passive plausible sentences (20) (Experiment 1). However, in sentences that were passive and implausible (21), individuals answered *the dog* 74% of the time, despite the argument structure of the sentence directly pertaining to the man as the agent and the dog as the patient (the man biting the dog) (Ferreira et al., 2002; MacDonald, 1994; Meng & Bader, 2021).

19) The dog bit the man.

20)The man bit the dog.

21) The man was bitten by the dog.

22) The dog was bitten by the man.

In Experiment 2 and 3, Ferreira (2003) examined sentence structure and the reversibility of sentences in addition to the plausibility of sentences. Latency and accuracy were measured in self-paced reading tasks where participants were asked to state what the agent or patient did (e.g., '*Who did the biting?*'). The complexity of sentences was manipulated by presenting sentences that could be reversed (23), sentences that were non-reversible (24) and sentences that were symmetrical (25) in plausible and implausible contexts. Furthermore, these sentences were presented where the sentence either included a subject cleft (Experiment 2) or an object cleft (Experiment 3). Sentences with a subject-cleft ('*The dog bit the man*.') were amended to an object cleft ('*It was the dog that bit the man*.') with minimal changes to the text.

23) The dog bit the man. (Subject cleft, reversible)

24) The mouse ate the cheese (Subject cleft, non-reversible)

25) The woman visited the man. (Subject cleft, symmetrical)

Similarly to findings between active and passive sentences, sentences containing object clefts were found to be more difficult to understand. Ferreira advocated that the results from these experiments together suggest that readers prefer sentence types where *heuristics* can be heavily relied upon in sentence processing, especially the *noun-verb-noun heuristic*, which is more difficult to use in passive and object cleft sentences. Schematic knowledge would suggest that the more plausible interpretation is the dog biting the man or the agent as the dog and the man as the patient. However, heuristics alone cannot explain sentence processing, instead Ferreira advocated that *fast and frugal* heuristics are used initially followed by algorithmic processing (Ferreira, 2003; Todd & Gigerenzer, 2000).

Despite these results having been replicated (Christianson et al., 2010; Lim & Christianson, 2013; Zhou & Christianson, 2016), some studies have found no differences in accuracy between active and passive sentences (Bader & Meng, 2018; Grillo et al., 2019; Meng & Bader, 2021), and more recently canonical and non-canonical sentences were used to establish if representations processed by the human parsing mechanism (HPM) are algorithmically computed or not (Meng & Bader, 2021). Similarly to Ferreira (2003), sentences were presented in an active, subject before object order '*The teacher visited the gardener.*', object before subject order '*It was the gardener who visited the teacher.*' and a passive structure '*The gardener was visited by the teacher.*'. Interestingly, task dependent results were found with similar results to Ferreira (2003) when a question probe task was used. However, when the task involved plausibility judgements, fewer errors were found. The results did not show support for passive sentences resulting in more parsing errors than canonical sentences and this was taken as evidence against good enough processing. The authors suggested that Ferreira’s (2003) study reflects task-specific retrieval processes. This role of shallow processing in pragmatic normalisation has been investigated more widely. For instance Paape et al. (2020) investigated whether depth charge illusions are processed incorrectly because of shallow processing. Participants were presented with control ‘*Some detail is too important to get left out*’ and depth charge sentences ‘*No detail is too unimportant to get left out*’. Context was manipulated by presenting participants with different instructions: a) participants read sentences in the context of the ‘normal world’ or b) participants read sentences and had to mentally invert the interpretation of sentences. To aid participants following task instructions, sentences in the ‘normal world’ were presented in black text on a white background. For the negation context, white text was presented on a black background. Results found a difference between control sentences, in the normal world sentences were considered to be more acceptable (86%) than in the negation context (32%). However, there was little difference for depth charge sentences in the normal (47%) or negation context (38%) (Experiment 4). The authors suggested that the results are consistent with the shallow processing framework as no difference in reading times were found for depth charge sentences suggesting that the sentence went underspecified. However, the findings could also be due to overloading information in working memory (Kizach, 2015; Paape, 2021; Wason & Reich, 1979).

An additional factor in the processing of passive sentences is individual differences. Research on this area is limited, however, an interesting study investigating individual differences in comprehension of passive sentences by native and non-native English speakers used a modified version of sentences presented in Ferreira (2003) (Dabrowska & Street, 2006). Results found highly educated native and non-natives, and less-educated non-native adults performed at ceiling for all conditions. In contrast, less-educated native adults had more difficulty in identifying '*Who the do-er is?*' in implausible passive sentences (36%) than implausible active sentences (65%). The results replicate Ferreira's (2003) findings, further suggesting good enough or shallow processing. Furthermore, the authors suggested that metalinguistic skills may affect processing of implausible passive sentences, hence why non-native and highly-educated adults had a higher proportion of correct answers. This experiment did not control for IQ or working memory, thus other individual factors affecting processing are not well understood.

There has been limited research into age differences, specifically, differences between younger and older adults' processing of plausibility and passive sentences and no eye-tracking study to date has been conducted. Chapter 6 introduces literature on aging and processing of passive sentences.

2.4. Shallow processing and semantic ambiguity (discourse level)

So far, this chapter has focused on when shallow processing results at the word- and sentence-level. However, readers may also engage in shallow processing at the wider discourse level. For example, research has found when readers encounter words that have similar meanings (such as 'newspaper' which can refer to the printing company or the paper which news is printed on), no one interpretation is immediately adopted (Frazier & Rayner, 1990). In contrast, when a word (e.g. bank) has very distinct meanings (for instance, 'bank' can describe the financial establishment or an area of land which slopes down), readers immediately commit to an interpretation (Frazier & Rayner, 1990). Underspecification does not occur in this instance because readers are able to align the context and features from the sentence to form a distinct meaning. On the other hand, when a word has similar meanings, then no one representation is forced to dominate and as a result, readers are likely to engage in underspecification. The influence of context on processing of semantic ambiguity has been explored widely, especially the processing of anaphora. Poesio et al. (2006) conducted a corpus study where participants read corpus transcripts between two individuals, one of which was a railway manager and the other individual was tasked with providing information on timetabling, equipment and any other required information. Eighteen participants were required to go through the transcripts and mark anaphors to the antecedent in what participants assumed was a simple annotation task (9) (Poesio et al., 2006). Half of the markers scored (e) as ambiguous. Of the nine markers who found the first pronoun (it) to be unambiguous, three thought *it* referred to the boxcar and six marked *it* as belonging to engine E2. Further ambiguity occurred with the second pronoun in (e.) (see Table 2).

Table 2

Example sentences from Poesio et al., (2006)

Line label	Character	Transcript
a.	M:	first thing I'd like you to do
b.		is send engine E2 off with a boxcar to Corning to pick up oranges
c.		Uh as soon as possible
d.	S:	Okay [6s]
e.	M:	And while it's there it should pick up the tanker

Semantic ambiguity can also occur when two or more quantifiers are included in a sentence (e.g. 21). Quantifiers are words that convey the number of entities (e.g., *every*, *each*, *all*, *a*). There are several theories that aim to describe what representation occurs in doubly quantified sentences (see Chapter 7 for more detail about quantifier theories).

Numerous theories have aimed to explain how the language system resolves semantic ambiguity when two or more quantifiers are in a sentence (Ioup, 1975; Kurtzman & Macdonald, 1993; Lakoff, 1971). In their influential article, Sanford and Sturt (2002) used sentences like (26) to illustrate how readers engage in shallow processing when quantifiers (*a* and *every*) are presented in doubly quantified sentences. Sentence (26) is semantically ambiguous as the sentence could refer to Kelly showing lots of photos to one critic in which case, (26a) would be easier to process. Alternatively, the interpretation of (26) could be that Kelly should lots of photos to more than one critic and in this instance, sentence (26b) would be easier to read. Interestingly, no differences in processing (26a) or (26b) were found suggesting that the representation of photos and critics were underspecified by the human processor (Tunstall, 1998).

26) Kelly showed every photo to a critic last month

26a) The critic was from a major art gallery

26b) The critics were from a major art gallery

More recently, research has been conducted that attempts to manipulate context to better understand the role context has in quantifier scope processing (Dwivedi, 2013). In three self-paced reading experiments, Dwivedi manipulated the context of doubly quantified semantically ambiguous sentences. In Experiment 1, Dwivedi presented ambiguous and unambiguous doubly quantified sentences followed by a singular or plural continuation sentence (see Table 3). In Experiment 2 the same sentences were presented with the addition of explicit questions querying the number of entities (e.g. *trees*). In Experiment 3, doubly quantified ambiguous sentences were presented that were not biased to a plural representation (which was the case in Experiment 1 and 2).

Table 3

Experimental sentences presented in Dwivedi et al., (2013)

	Context	
Continuation	Ambiguous	Unambiguous
Plural	Every kid climbed a tree. The trees were in the park.	Every kid climbed those trees. The trees were in the park.
Singular	Every kid climbed a tree. The tree was in the park.	Every kid climbed that tree. The tree was in the park.

In Experiment 1, reading times for ambiguous sentences were faster than unambiguous sentences indicating these sentences were underspecified. No additional effects were found on the continuation sentence (e.g. *The tree was in the park*). Findings from Experiment 2 showed that it was only when questions were presented that computation took place with a preference for the plural interpretation. These findings were also found with unbiased ambiguous doubly quantified sentences (in Experiment 3). Dwivedi (2013) suggested scope computation does not take place unless it is demanded. Thus, it was advocated that these results were evidence of two types of processing taking place: firstly, heuristics and then algorithmic processing.

In conclusion, this chapter has shown that semantic ambiguities are often underspecified, especially in quantifier scope processing. If readers engage in shallow processing, doubly quantified sentences may be partially processed or underspecified. In this thesis, the interest is whether context can influence the processing of quantifier scope processing in young and older adults. Research has attempted to investigate this using a self-paced reading task (Kemtes & Kemper, 1999a). Thus, chapter 7 details an eye-tracking experiment investigating the role of context in the processing of semantic ambiguities using eye-tracking evidence from young and older adults.

2.5. Eye movement during online processing and offline comprehension

An aim of this thesis is to explore the eye movements of young and older adults during the online processing of text and relationship of this with good enough processing. There has been growing interest in understanding the underlying processes of language comprehension (Ferreira et al., 2019). Understanding the link between eye movements during reading and offline comprehension is crucial and has implications for models of language comprehension.

Previously, one study has examined the relationship between shallow processing and eye movement behaviour. Wonnacott et al. (2016) measured the relationship between eye movement behaviour and whether this predicts shallow processing at the sentence-level using temporary syntactic ambiguities. In this study, children and young adults' eye movements were recorded. The number of regressions made into earlier regions of text positively predicted comprehension, however, note that this was the case for both ambiguous and unambiguous conditions. Interestingly, Wonnacott et al. (2016) found no relationship between total reading times and shallow processing.

A further study has measured the eye movements of readers and examined the relationship of this with accuracy scores. In an eye-tracking study, Huang and Ferreira (2021) presented young adults with sentences containing temporarily syntactic ambiguities and found that longer first-pass and total reading times predicted more errors in comprehending garden-path sentences. No effect was found for go-past time.

2.6. Summary of shallow processing

This chapter has focused on describing key findings of shallow processing at the word-, sentence- and wider discourse-level, including several factors that influence shallow processing (see Sanford and Graesser (2006) and Christianson (2016) for special issues on this topic). An area that has received less attention is age differences in shallow processing. Older adults are of particular interest because over one's lifetime, elements of crystallised intelligence, such as, more experience and knowledge of the world is accumulated and vocabulary size expands. In contrast, elements of fluid intelligence, such as, working memory which is crucial for recall and integration of information during language processing deteriorate. Coupled together, older adults may be expected to engage in shallow processing more so than their younger counterparts to compensate for deterioration of elements of fluid intelligence.

Currently, there is evidence that shows older adult employ different strategies when reading. For instance, older adults have been found to employ a risky reading strategy where older adults are more likely to make longer forward eye movements (McGowan et al., 2015; McGowan et al., 2014; Paterson et al., 2013a; Rayner et al., 2006, 2011, 2013), skip words more frequently (McGowan et al., 2014, 2015; Paterson et al., 2013; Rayner, et al., 2006, 2011, 2013) but then regress back to words more often when their inferences are incorrect (Kliegl et al., 2004; McGowan et al., 2014, 2015; Paterson et al., 2013; Rayner et al., 2006, 2011, 2013). It has been suggested that older adults try to guess upcoming words to compensate for cognitive decline during aging but then regress more when inferences are incorrect.

Evidence of age differences in shallow processing have been most widely conducted at the sentence-level using temporary syntactic ambiguities. Some research has found older adults are more likely to engage in good enough processing (Christianson et al., 2006; Dede, 2014; Kemper et al., 2004). Age differences in shallow processing at the word and wider discourse level have been researched les (Dwivedi et al., 2010; Hannon & Daneman, 2004; Kemtes & Kemper, 1999a). The majority of these studies have focused on the proportion of correct answers to examine shallow processing and what is less understood is the age differences during online processing of sentences that may result in a shallow interpretation.

The studies within this thesis utilise eye-tracking to record the eye movements of readers to examine whether there is evidence of shallow processing online and whether age differences between younger and older adults occur. Additionally, the influence of working memory and vocabulary on shallow processing are examined. Finally, the relationship between eye movements and offline comprehension are explored. The next section focuses on the methodological and analytical approach taken for the experiments described in this thesis.

Chapter 3: Methodology and data analysis

Shallow processing has been examined using a range of different online experimental methods, most notably, self-paced reading tasks (Barton & Sanford, 1993; Christianson et al., 2006; Kemtes & Kemper, 1999) and more recently, eye-tracking experiments (Bohan & Sanford, 2008; Filik et al., 2004; Jacob & Felser, 2016) and Event Related Potential (ERP) studies (Bohan et al., 2012; Dwivedi & Gibson, 2017). The literature described in this thesis predominantly utilised self-paced reading methods (Barton & Sanford, 1993; Christianson et al., 2006; Meredyth et al., 2006; Ferreira, 2003; Kemtes & Kemper, 1999; Qian et al., 2018). In the following section, I describe the different methods and advantages of using eye-tracking to investigate shallow processing. Additionally, R was used to analyse eye-tracking data using General Linear Mixed Models, the approach taken is described in section (3.4).

3.1. Self-paced reading tasks

In the 1970s, psycholinguists created the self-paced reading task (Aaronson, 1976; Jegerski, 2014; Mitchell & Green, 1978). This task requires an individual to read text presented on a computer screen. A fixation point is initially displayed and the participant presses a button when they are ready to see the next word or phrase (often referred to as a region). After reading the region, the participant can press a button to unmask the next portion of text, which replaces the previous segment until completion of the trial. The reaction time of each portion of text is recorded in (ms) to establish processing difficulty. The measure of time includes the time taken to read a region as well as the time taken to press the button to move on to the next region or item.

There are many variations as to how text is presented in self-paced reading tasks (Jegerski, 2014). The **cumulative** display comprises of each portion of text remaining on screen (Ferreira & Henderson, 1991). In contrast, the **non-cumulative** display (Lim & Christianson, 2013) masks all text that is not the fixated region. Text can also appear in a linear or centred presentation (Jegerski, 2014). The most common self-paced reading method is the moving window paradigm (see Figure 7), where the entire item is presented on the screen but all words except the target region are masked by a series of dashes (punctuation remains unmasked) (Dwivedi, 2013; Just et al., 1982; Tan et al., 2020; Traxler et al., 2014). The first button press triggers the first set of dashes to be replaced by the displayed region. Subsequent button presses replace the preceding region with dashes and unmask the next region and so on. Comprehension questions are often incorporated into a subset of trials to ensure the participant is attending to the stimuli along with filler trials. Filler trials are non-experimental sentences.

Figure 7

Example stimuli in a moving window paradigm

```
The quick brown fox jumps over the lazy dog.

The xxxxx xxxxx xxx xxxxx xxxx xxx xxxx xxx.

Xxx quick xxxxx xxx xxxxx xxxx xxx xxxx xxx.

Xxx xxxxx brown xxx xxxxx xxxx xxx xxxx xxx.
```

The self-paced reading task is underpinned by two assumptions: a) the eye-mind hypothesis which suggests that the brain processes the word the eye fixates on (Just & Carpenter, 1980) and b) the reader processes the word encountered as early as possible as opposed to reading the whole sentence and then processing the meaning (Haberlandt, 1994). Studies involving self-paced reading tasks have provided a wealth of knowledge into shallow processing. Studies have shown that young and older adults have greater processing difficulty when reading temporarily ambiguous sentences, often resulting in shallow processing (Christianson et al., 2006; Qian et al., 2018). Furthermore, young adults are susceptible to shallow processing when plausible sentences are presented in a passive structure (Ferreira, 2003). At the word-level, semantic anomalies were found to be processed in a shallow manner and longer reading times were found when resolving a semantic anomaly that was highly related to the non-anomalous word (Van Oostendorp & De Mul, 1990). Though self-paced reading tasks are valuable, there are limitations to this method. For instance, in Van Oostendorp and De Mul's (1990) study, participants were tasked with reading sentences containing semantic anomalies and pressing a separate button for whether the sentence was considered ‘true’ or ‘false’. Longer processing times were found for statements considered to be true than false. The authors were unable to know whether longer reading times were found for true statements or if the time taken to press a key with the left hand accounted for this finding. Further limitations are discussed below.

3.2. Event Related Potentials

ERPs are an average measure of a fixed or locked recording of electrical changes in the brain. ERPs are measured on a millisecond-to-millisecond timeframe to highlight the cognitive and perceptual changes occurring during language processing over numerous trials. ERPs are commonly calculated from the raw electroencephalogram (EEG) (Mecklinger et al., 1995; Sereno & Rayner, 2003). The number of electrodes used in studies varies but commonly a high-density recording of the scalp is conducted eliciting spatial and temporal information (Mecklinger et al., 1995; Sereno & Rayner, 2003). ERP wavelengths can be short (around 300ms) or long waveforms (over 1000ms).

During typical ERP studies, words are presented in a Rapid Serial Visual Presentation (RSVP) to minimise ocular movements which EEG recordings are sensitive to (Degno et al., 2021). This involves participants reading sentences where words are presented one word at a time in a centered format, the duration of the presentation of one region to another is called stimulus onset asynchronies (SOAs) and can vary anywhere between 400ms to 1000ms (Dambacher et al., 2012).

Over decades, numerous ERP components have been identified which reflect the neural activity underlying language processing. Specifically, to this thesis, difficulties in syntactic processing, such as, ‘The friend was in the visited’ have been shown to commonly elicit a parietal P600 effect compared to control sentences presented in the study, such as, ‘The finder was rewarded’ (Friederici et al., 1993; Osterhout & Holcomb, 1993). In contrast, semantic violations, such as, ‘He spread the warm bread with socks’ elicit an N400 centro-parietal negativity 400ms after the onset of the semantically anomalous word compared to control sentences such as ‘He spread the warm bread with butter’ (Kutas & Hillyard 1980; Kutas & Federmeier, 2011; Seyednozadi et al., 2021).

Context has also been found to influence ERPs, for instance, in sentences like, ‘Wang walked through a narrow street and then turned into a yard’, a larger N400 amplitude is identified, however when sentences includes a more predictable context, such as, ‘Wang heard a sharp sound and screamed out immediately’ (Jiang & Zhou, 2020; Kutas & Hillyard, 1980; Zunini et al., 2020) this is not found. Interestingly, in some studies an N400 component was expected to be evoked but instead a P600 component was elicited. For instance, in thematically incongruent sentences like ‘For breakfast the eggs would only eat…’ and control sentences like ‘For breakfast the boys would only eat…’ an N400 was expected yet a P600 component was elicited. It may be that the P600 reflects conflict between the syntactic and semantic processing and processing of plausibility may be delayed (Sikos et al., 2016). These studies illustrate the that when processing complex sentences both the P600 and N400 are elicited.

Many of the elements related to sentence processing have been found to be stable across the lifespan and in some cases, such as vocabulary knowledge strengthens across the lifespan (Federmeier & Kutas, 2019; Salthouse, 1990). Older adults are able to utilise their experience and acquired knowledge of the world to ease processing of sentences and compensate for any cognitive declines (Salthouse, 2012). ERP studies have found a number of age-related differences. Older adults are less able to accumulate contextual information as lower activation in the N400 has been found (Payne & Federmeier, 2019). Studies have shown older adults are less likely to elicit an N400 component for incongruous semantic violations (Wlotko & Federmeier, 2012). In semantic violations, young adults have asymmetric processing to the left hemisphere, which is associated with language and is the area that N400 activation occurs. However, older adults were not found to have asymmetric hemisphere processing and no initiation in left or right hemisphere was identified for semantic prediction (Federmeier & Kutas, 2019). Compared to self-paced reading tasks, ERP studies highlight the complex nature of sentence processing and that processing is far from complete at a single time point. Furthermore, ERP studies have effectively shown that there are differences in sentence processing due to aging.

3.3. Eye-tracking

The ability to track a reader's eye movements has advanced considerably. Louis Emile Javal's observation of eye movements where important information was realised, such as fixation durations and saccade latency instigated the recording of eye movements (Huey, 1908). From recording the naked eye, advances in technology have transformed our understanding of eye movements during reading (Rayner, 1978, 1998, 2009). Initially, eye-tracking was an invasive and uncomfortable method as it involved measuring the skin around the eye using electrodes to measure any differences in electric potentials (Singh & Singh, 2012). Another method required participants to wear contact lenses large enough to cover both the cornea and sclera. This contact lens had a metal coil attached to it, which moved when the eye moved, allowing for eye movements to be measured by electromagnetic field changes (Duchowski, 2003; Płużyczka, 2018). Current eye-tracking methods utilise both infrared light and Information Technology. In typical eye-tracking methods, an infrared camera sits underneath a standard desktop computer, this camera is linked to eye-tracking software on a separate computer, which records the participant's eye movements. This is possible because of LED light (from the infrared camera) that is directed to the participant's eye, entering the pupil creating a reflection of the cornea. The imaging software creates a vector between the corneal reflection and the pupil, using specialised formulas (for further details see Hansen & Ji, 2010; Poole & Ball, 2005). For an accurate measurement, a calibration process takes place. A calibration involves a small dot appearing on different areas of the screen. As the participant's eye moves to the location of the dot, x-y coordinates are created mapping to the pupil-centre/corneal reflection relationship. Calibration can involve a 3 to 13 grid pattern and is normally repeated multiple times throughout the experiment (Goldberg & Wichansky, 2003). Usually, only the dominant eye is monitored during reading experiments. Though there is some incongruity between the location of the fixations of the left and right eye, however, no age differences have been identified (Paterson et al., 2013a).

In a typical eye-tracking experiment, a participant reads text on a computer screen with their chin and forehead on rests to minimise any head movements. The presentation of text is somewhat natural in that whole sentence(s) are presented and participants read silently and at their own pace. To move onto the next trial, a device, such as a game pad button is pressed. Similarly to self-paced reading tasks, a fixation point is presented before each trial. Whilst the participant reads, the location, duration and sequence of the eye movements are recorded (Frazier & Rayner, 1990). There are two main eye movements: saccades and fixations. Saccades are jerky, ballistic movements and allow the reader's eyes to move to different locations across the screen or page (Rayner, 1978). Fixations are brief pauses in which the reader can obtain visual information (Rayner, 1978). Blinks are also recorded but these are normally discarded as they do not provide useful information. Eye-tracking allows for numerous eye movement measures to be calculated.

Eye movement measures are calculated and are often referred to as 'early' or 'late' measures. These terms do not accurately reflect eye-movement models, instead, 'early' and 'late' measures often denote information about underlying processes (Clifton et al., 2007; Vasishth et al., 2013). Early measures often relate to where the reader has first come into contact with the region and late measures denote the reader has reanalysed the area. For instance, first-pass reading times are categorised as an 'early' measure as first-pass reading effects found in a region signal the immediate difficulty of the manipulation for readers (Liversedge et al., 1998). First-fixation duration, single-fixation duration, regressions out of a region and skipping of a region also reflect 'early' processing (see Table 4). Regression path reading times are classed a 'late' effect as this measure often indicates that the effects of the manipulation have occurred later in processing, such as integrating information together (Liversedge et al., 1998). The number of regressions made in to a region and total reading times indicate 'late' processing (see Table 4). In this thesis, effects are expected to occur in both 'early' and 'late' measures for young and older adults.

Table 4

Common eye movement measures reported in eye-tracking studies

Measure	Definition
First-fixation duration	The duration of the first forward fixation on a region of interest.
First-pass reading time / Gaze duration	The sum of all fixation durations from the eyes first entering the specific region to first leaving it. The difference between gaze duration and first-pass is that gaze duration is used when the region is a single word.
Total number of fixations	The total number of fixations made in a region.
Regression path reading times / go past	Time from the eyes first fixating on a given region plus all time spent fixating to the left of the region before fixating on a subsequent region to the right for the first time.

Probability of skipping	Indicates the number of trials where an area of text is not fixated during first-pass reading.
Regressions in	Reflects the proportion of trials where a saccade is made back (regressive saccade) into a region.
Regressions out	Is described as the proportion of trials where a saccade is made out of the region during first-pass reading.
Total reading times	Is the sum of all fixations a reader makes in a given region.
Second-pass reading time	Represents the total duration of rereading (all time excluding first-pass reading times).

Eye-tracking methods have a number of advantages compared to self-paced reading tasks and ERP studies, the most obvious being that the reader is able to participate in natural reading. Self-paced reading is limited to the reader pressing a key after each region which does not reflect the natural reading process. Furthermore, ERP studies are prone to artificial sentence presentation as one word is presented for a set time. Eye-tracking studies are more ecologically valid, allowing the participant to read an entire sentence or paragraph of text and the button press is only used to move on to the next trial.

Secondly, self-paced reading tasks are limited to recording time spent on a given region but cannot measure the various types of eye movements made. Typically, readers look back at earlier parts of text up to 15% of the time and skip some letters or short words all together (Rayner et al., 2006). Moreover, eye movements that occur in natural reading, such as, regressions and word skipping are eliminated in an ERP experiment and this has been found to induce a different reading behaviour (Kliegl et al., 2012; Metzner et al., 2017;

Rayner, Reichle, et al., 1998; Schotter et al., 2014). Eye-tracking enables the tracking of the moment-to-moment reading process allowing for the calculation of a number of measures such as regressions made in to and out of regions, skipping behaviour as well as fixation times. Though self-paced reading tasks have commonly been used to study the processing of ambiguous or anomalous sentences, these types of sentences incur costs in both early and late processing often evidenced by first-fixation, first-pass, regression path and total reading times (Liversedge et al., 1998), which may not be reflected in the limited measures of self-paced reading.

Finally, eye-tracking is less invasive than ERP studies, making eye-tracking well suited to the older adult sample used in the experimental studies in this thesis. Similar to ERP and self-paced reading studies, eye-tracking experiments usually include a number of comprehension questions which are presented after a proportion of trials. This aim of this thesis is to investigate the age differences in the depth of processing at the word-, sentence- and discourse-level. A secondary aim is to examine the relationship between eye movement measures and shallow processing. To do this, it was deemed important to understand differences in both early and late processing to fully examine shallow processing. These questions cannot be effectively answered by self-paced reading tasks or ERP studies. Accordingly, eye-tracking was considered the most detailed and appropriate method for the experiments designed for this thesis.

3.4. Data analysis

Traditional methods of data analysis for eye-tracking studies have primarily used F1 (participant) and F2 (item) ANOVA analyses, this involves averaging by participants and items separately and then by condition (Raajmakers, 2003). There are numerous issues with this method. For example, the items are randomised but participants see the same set of items, which results in the mean effect in items and participants being different. There are also issues in interpretation in that ideally there should be significant effects by both participant and item. However, often a significant effect is found for only one F statistic, which results in ambiguity when interpreting results.

Issues related to analysing language experiments have recently been mitigated by using Linear-Mixed-Effects-Models (LMEM) in R (Baayen et al., 2008). In addition to being a more powerful and flexible approach (Locker et al., 2007), LMEM enables the simultaneous account of error variance for both participant and item in the same statistical model (Baayen et al., 2008). Furthermore, due to the power of LMEM, it can withstand violations of homoscedasticity and sphericity (Quene & Van Den Bergh, 2004), using all data points instead of the averages used in ANOVA. LMEM differs to the traditional linear regressions in that it includes random effects which accounts for variance in random sampling, such that any intrinsic factors (i.e., some items are interpreted easily or variability amongst participants) are accounted for with the inclusion of random intercepts and slopes, accounting for any random error that may affect results.

There is no standardised method to running LMEM. In published papers, Barr et al. (2013) are often credited with describing a structure in which to run LMEM for a psycholinguistic study. The common method is to run a full 'maximal model', including both random intercepts and slopes to avoid being anti-conservative, meaning to underestimate the variation of variables. An example of a maximal model for language experiment is: Outcome variable ~ predicted variable 1*predicted variable 2 + (1 + predicted variable 1*predicted variable 2 | Participant) + (1 + predicted variable 1*predicted variable 2 | Item). However, LMEM can be problematic in that the models are prone to non-convergence meaning that there are issues with the fit of the model. Issues are often due to the power or the structure being too complex in which case the model should be trimmed beginning with the random effects structure. Alternatively, the iterations can be increased until model convergence occurs (Barr et al., 2013) by using an optimizer, in this thesis “bobyqa”. Models can be compared using the ANOVA function. The output also includes the model’s Akaike information criterion (AIC) value and Bayesian information criterion (BIC) value which can be used to determine the best fitting model. Despite the issues described with LMEM, most published studies now use LMEM because of the multiple advantages that this method involves.

Prior to running LMEM, evidence of normality, linearity and homoscedasticity is inspected using the raw (untransformed) residual plots. A common issue is that time-series data is often positively skewed, especially in studies involving older adults (Lo & Andrews, 2015). To mitigate any deviations, usually log transformation of data is conducted to account for skew of time series data, however this renders the results difficult to interpret as it distorts the ratio scale, which creates issues similar to the original F1 and F2 ANOVA analysis. An approach to overcome this is to use the GLMM model which unlike LMEM relaxes the assumption of a normal distribution. Specifically, a gamma family and identity link are applied following Lo and Andrews (2015). These links define the distribution as Gamma and similarly to LMEM that the relationship between the predictor variables and the outcome variables is directly linked or linear. The lme4 package is used to run GLMM (and LMEM) in *R* (Bates et al., 2015). The nature of testing two samples where older adults have been found to have longer reading times creates an heteroscedastic pattern by nature. The GLMM was considered the most appropriate approach for analyses of the experiments presented in this thesis as it results in a more powerful adjustment for slower reading times, thus greater reliability when assessing individual differences. Another important consideration in data analysis of psycholinguistic studies is the handling of outliers (Jones, 2019; Osborne & Overbay, 2004). A common approach is to implement a cut off, this can be between 2-3 SD.

A 2.5 SD from the mean was used per participant, per condition for each dependent variable. Despite Baayen (2008) being cited over 7000 times, the majority of the studies described in this thesis have not used LMM and to my knowledge none have used GLM and so this is still a fairly novel approach.

For each of the experiments described in this thesis, a maximal model was initially conducted e.g., in experiment 2 the maximal model for total reading times would be Total reading times ~ Length*Ambiguity*Age + (1 + Length*Ambiguity*Age | Participant) + (1 + Length*Ambiguity*Age | Item). If a model failed to converge, as well as increasing the number of iterations and using the optimizer “bobyqa”, the model was trimmed using the technique described by Barr et al. (2013) - removing correlations, interactions, and slopes until convergence was successful. *P* values were automatically calculated from the lmerTest package (Kuznetsova et al., 2017). Only results from successful model convergence are reported (please see the GLMM table statistics for each chapter). Working memory and vocabulary were centered and added as covariates: Total reading times ~ Length*Ambiguity* Age + Working memory/Vocabulary + (1 + Length*Ambiguity*Age | Participant) + (1 + Length*Ambiguity*Age | Item). If the model converged, then the model was compared to the base model using the ANOVA function. Models including working memory or vocabulary often did not converge or if they did then they were not deemed a better model. If an interaction was statistically significant, then simple effects were conducted for interactions identified in the GLMM. The fixed factors were combined to create a new factor with the necessary levels e.g., for a length*ambiguity interaction: long_ambiguous, long_unambiguous, short_ambiguous and short_unambiguous sentences. Contrasts were created e.g., c(1,-1,0,0) and a GLMM model was conducted on this combined factor e.g., Total reading times ~ combined + (1 + Length*Ambiguity*Age | Participant) + (1 + Length*Ambiguity*Age | Item). To examine the relationship between accuracy and eye movement measures, a logistic regression was conducted e.g., accuracy ~ eye movement measure.

3.5. Overview of thesis

In this thesis, I aim to utilise eye-tracking technology to record the eye movements of young and older readers to examine whether they engage in shallow processing online. Four eye-tracking experiments were conducted to examine whether age differences in shallow processing between younger and older adults occur at the word-, sentence-, and wider-discourse level. Additionally, the relationship between eye movements during reading and offline comprehension was explored. Working memory and vocabulary assessments were conducted to understand how these factors may modulate shallow processing in young and older adults.

3.5.1. Experiment 1

In Chapter 4, Experiment 1 comprised of examining age differences in the processing of sentences at the word level. Eye movements were monitored for 32 young adults (aged 18-30 years) and 32 older adults (aged 65+ years) while they read sentences containing hard-to-detect semantic anomalies (e.g., *The authorities negotiated / communicated with the scared and desperate hostages and calmed them down.*). A manipulation was included which rendered the critical region (*hostages*) either anomalous (i.e., when preceded by *negotiated*) or non-anomalous (i.e., when preceded by *communicated*). In addition to detection, evidence of 'unconscious registration' of undetected anomalies was also investigated as the shallow processing framework would advocate that unconscious registration would not occur if anomalies go undetected. A further aim focused on exploring the relationship between eye movement behaviour during reading and detection accuracy. Finally, working memory and vocabulary assessments were conducted for young and older adults using the WAIS-IV.

3.5.2. Experiment 2

Previous research suggests adult age differences exist in the processing and comprehension of temporary syntactic ambiguities during reading. In Chapter 5, an eye movement experiment was described which compared the underlying cognitive processes when young (18-30 years old) and older adults (65+) read temporarily ambiguous sentences, such as, “While the gentleman was eating the burgers were still being reheated in the microwave.” Participants were then asked comprehension questions to assess depth of processing. Additionally, the influence of working memory was investigated by manipulating the amount of text between the ambiguity and subsequent disambiguating information, and by assessing working memory capabilities for both age groups. Similarly to Experiment 1, the relationship between eye movements and accuracy of comprehension questions was explored. Performance on working memory and vocabulary assessments were recorded for young and older adults.

3.5.3. Experiment 3

The aim of the study in Chapter 6 was to further examine whether there is a relationship between eye movement behaviour during reading and good enough processing in both young and older adults, in order to further our understanding of the relationship between online processing and ultimate interpretation of text. Eye movements were monitored for 32 young adults (aged 18-30 years) and 32 older adults (aged 65+ years) while they read active or passive sentences describing either plausible or implausible events. Analyses focused on replicating Ferreira’s (2003) study, examining whether older adults engaged in risky reading and the relationship between eye movement characteristics of risky reading and accuracy (which was used as an index of good enough processing).

3.5.4. Experiment 4

The aim of Experiment 4 in Chapter 7 was to examine shallow processing at the discourse-level. Doubly quantified sentences (e.g., *Every kid climbed a tree*), are ambiguous regarding the number of entities that are represented by the indefinite phrase (*a tree*). Previous research has investigated the influence of both structural (i.e., the order in which the quantifiers appear) and semantic factors (i.e., properties of the quantifiers themselves) on the on-line processing of subsequent anaphoric reference to the indefinite phrase (e.g., *The tree was* vs. *The trees were*). Non-linguistic factors have received much less attention. In Chapter 7, the influence of context and individual differences, specifically, age and working memory capacity were examined. Eye movements were monitored for 48 young adults (aged 18-30 years) and 48 older adults (aged 65+ years) while they read sentences with a contextual bias for a singular entity (e.g., *Every student in the class is listed on a register*) or plural entities (e.g., *Every school pupil in the country has their attendance marked on a register*). A singular (*This register…*) or plural (*These registers…*) continuation then followed. Participants' working memory capacity was also assessed.

Chapter 4: Anomaly detection

4.1. Introduction

When asked, *When an airplane crashes, where should the survivors be buried?*, individuals often fail to notice that survivors are those who are alive, and therefore do not need to be buried (Barton & Sanford, 1993; Patson & Warren, 2010). Despite knowledge of the meaning of *survivors*, readers often fail to detect such semantic anomalies. This "semantic illusion" is an example of shallow processing as it highlights the way in which readers often do not process all of the semantic information in sentences, instead relying on the context and their world experience which leads to a shallow representation (Sanford & Sturt, 2002). Decades of research into anomaly detection has shown that individuals consistently fail to detect semantic illusions (Barton & Sanford, 1993; Bredart & Modolo, 1988; Erickson & Mattson, 1981; Hannon & Daneman, 2001, 2004; Van Oostendorp & De Mul, 1990), especially when the anomalous word has a good global fit to the context (Bohan & Sanford, 2008; Erickson & Mattson, 1981; Sanford & Sturt, 2002).

Relatively little is known about individual differences in anomaly detection, or the cognitive mechanisms involved. The aim of this chapter is to investigate the influence of individual differences, specifically age, on the detection (and non-detection) of hard-to-detect anomalies using eye-tracking during reading to examine moment-to-moment processing. The processing of anomalies that go undetected is of particular interest, since if online processing of undetected anomalies is similar to that of detected anomalies, it indicates that undetected anomalies may be nevertheless unconsciously registered – an idea which is currently debated in the literature. It is also of vital importance to examine the relationship between eye movement behaviour and comprehension accuracy; an area which has been somewhat overlooked (Ferreira et al., 2019), but is fundamental to our understanding of sentence processing and reading comprehension.

Evidence for unconscious registration of undetected (i.e., unreported) semantic anomalies has been mixed. For example, an initial study conducted by Bohan and Sanford (2008) found no evidence for the online processing of undetected anomalies being any different to that of non-anomalous sentences. In their study, young adult readers' eye movements were monitored while they were reading semantically anomalous and non-anomalous sentences (e.g., Example 1). In Example 1, a context region was manipulated (*negotiated* / *communicated*) which rendered the critical region anomalous or non-anomalous (*hostages*).

1) The authorities / {negotiated (anomalous context) /communicated (non-anomalous context)} / with the scared and desperate / hostages (critical region) / and calmed them down.

In this example, *hostages* is anomalous when the context is *negotiated* because hostages are not the group to negotiate with in a hostage situation. The control manipulation is non-anomalous (e.g., *communicated*) as it is conceivable that hostages want information about their stressful situation. Participants were asked to notify the experimenter of anything 'out of place' in the sentences. In the critical region (*hostages*), no immediate reading difficulty was found, however, findings in later measures of reading behaviour indicated that greater reanalysis took place for detected anomalous sentences compared to undetected and non-anomalous sentences. Importantly, no differences between non-anomalous and undetected anomalous items were found, suggesting that anomalies did not result in additional processing, unless detected (i.e., there was no evidence of the anomaly being 'unconscious processed').

A follow up study by Sanford et al. (2011) used event-related potentials (ERPs) to examine whether undetected semantic anomalies elicited an N400 effect. This effect has been found to be elicited by easy-to-detect semantic anomalies (Van Berkum et al., 1999) and violations of world knowledge (Filik & Leuthold, 2008). Thus, if an N400 effect was found for undetected semantic anomalies compared to non-anomalous controls, this would be taken as support for unconscious registration of undetected anomalies despite them not being reported. However, in this study hard-to-detect anomalies were not found to produce an N400 effect compared to the non-anomalous condition. Furthermore, when easy-to-detect anomalies were presented an N400 effect was elicited compared to the non-anomalous condition (Sanford et al., 2011). These studies (Bohan & Sanford, 2008; Sanford et al., 2011) were taken as evidence that readers engage in shallow processing (Sanford & Sturt, 2002). shallow processing arises due to a fault in the retrieval process or integration of a word meaning with the wider discourse and as a result, the anomaly does not register in the reader's comprehension system.

Subsequent studies have reported contrasting results. For example, Cook et al. (2016) used eye-tracking to investigate the eye movements of young adults when reading sentences containing semantic anomalies. Specifically, true/false statements were presented to participants, such as, *Dogs usually don't like the sound of thunder during a storm*. Versions of the sentences could contain anomalous words that were either highly related (*lightning*) or highly unrelated (*sunshine*) to the non-anomalous information (*thunder*). Results showed that all anomalies incurred longer reading times when compared to non-anomalous sentences. However, only highly unrelated detected anomalies (e.g., *sunshine*) incurred longer initial processing times. Hannon (2015) found similar results, that is, anomalies that went unreported nevertheless generated longer reading times.

Cook et al. (2016) advocated that the RI-Val model (Cook & O'Brien, 2014; O'Brien & Cook, 2016) best accounts for these findings, rather than the shallow processing framework (Sanford & Sturt, 2002). There are three stages to the RI-Val model which explains online processing; the *Resonance* mechanism describes the process of incoming information being encoded with information stored in the long-term memory; *Integration* – in this stage, information processed in the resonance stage is integrated with information in the working memory; *Validation* – the interpretations derived are then validated using a low-level partial matching procedure. Processing is complete when the processing is considered 'good enough' via a *coherence threshold*. The RI-Val model suggests all anomalies are processed in the same manner, however Resonance, Integration, and Validation stages occur late for undetected anomalies, resulting in delayed processing difficulty. On the other hand, the shallow processing account suggests that anomalies go undetected despite processing taking place because the undetected anomaly does not register in the reader's comprehension system.

4.1.1. Aging

Of particular relevance to the present study are individual differences in semantic anomaly detection. Age differences are of interest because over one's lifetime more experience and world knowledge are acquired (Ben-David et al., 2015; Christianson et al., 2006; Keuleers et al., 2015). Anomalies are more likely to go undetected when the anomaly fits with the global context of the sentence. As a result of the semantic anomalies having a 'goodness of fit', readers may be less likely to retrieve the meaning of the word or integrate it with the wider discourse (Barton & Sanford, 1993). Following this, older adults may be more likely to rely on shallow processing due to the greater knowledge acquired during their lifetime and to compensate for any age-related cognitive decline.

To the authors' knowledge, only one study has examined semantic anomaly detection in older adults. Daneman et al. (2006a) used eye-tracking to assess the processing of easy and hard-to-detect semantic anomalies compared to non-anomalous sentences, testing young and older adults. Participants' accuracy in anomaly detection was also recorded. Results indicated that young adults' and older adults' accuracy scores were no different. However, older adults' eye movement data suggested that this group were more sensitive to semantic anomalies - at the anomalous region immediate difficulty was identified in older adults' eye movement data, as well as more time spent integrating the information. However, this study had a limited set of materials from which to draw these conclusions (two anomalous words, one non-anomalous, and four filler sentences) with many of the results not reaching significance. Thus, it is still largely unknown whether age differences occur in processing semantic anomalies.

4.1.2. Eye movement behaviour and depth of processing

Typically, studies of language comprehension will assess either the ultimate interpretation of a piece of text using measures of offline comprehension, such as questions to probe interpretation, or will examine the moment-to-moment processes during online reading behaviour. Increasingly, researchers are recognising the importance of studying the relationship between online processing behaviour and offline interpretation measures in order to gain a fuller understanding of the processes underlying language comprehension (e.g., Ferreira et al., 2019). Initial evidence suggests that the number of regressive eye movements made during reading may influence depth of processing (Wonnacott et al., 2016; Wotschack & Kliegl, 2013) and not being able to make regressions may lead to more errors when processing ambiguous sentences. However, relatively little is still known about this relationship, and in particular, whether it varies across the adult lifespan.

4.1.3. Present experiment

To establish whether undetected semantic anomalies are unconsciously registered despite not being consciously reported, an eye-tracking experiment was conducted to examine the eye movements of young and older adult readers while reading sentences containing hard-to-detect semantic anomalies. We were particularly interested in differences between young and older adult readers as older adults may be more inclined to not detect semantic anomalies if they are more likely to rely on their world knowledge and context. Sentences such as (2) were presented and the context manipulation rendered the critical region (*Glasgow*) anomalous (*capital city*) or non-anomalous (*largest city*).

2 The Scottish tourist board is trying to attract more tourists to Scotland who want short city breaks. One popular destination is also the largest/capital city of Scotland, which is Glasgow on the river Clyde.

Each participant was asked to read for comprehension and to identify any words that appear to be 'out of place'.

4.1.4. Predictions

The shallow processing framework (Sanford & Sturt, 2002) predicts greater difficulty during online processing for detected anomalies only. Specifically, following the predictions of this framework, we would expect longer reading times and more regressions to earlier parts of text for detected anomalies compared to non-anomalous and undetected items (Bohan & Sanford, 2008). In contrast, following the predictions of the RI-Val Model (Cook et al., 2016), if there is evidence of unconscious processing of undetected anomalies, the online processing of detected and undetected anomalies would be expected to be similar, but processing difficulty for undetected anomalies would be observed in later eye movement measures than for detected anomalies.

In terms of predictions related to eye movements and age, in line with the findings of Daneman et al. (2006a), older adults may identify anomalies earlier than young adults, thus, older adults would experience greater initial difficulty in processing. In addition, later eye movement measures may identify older adult readers being more inclined to fixate on the critical region for longer, producing longer total reading times than young adults, and older readers may look back to earlier parts of the text to try to integrate information together (Daneman et al., 2006a).

In terms of detection accuracy and age, predictions based on the aging literature would suggest that older adult readers may be more inclined to utilise their world knowledge and experience, relying more on shallow processing, and as a result may be less likely to correctly detect semantic anomalies than their younger counterparts. On the other hand, if our findings are similar to those of Daneman et al (2006a.), we would expect no age differences.

Little research has been conducted into the relationship between depth of processing and measures of eye movement behaviour during reading. However, one previous study has found that more regressions back to earlier regions of syntactically ambiguous sentences were linked to fewer errors on comprehension questions (Wonnacott et al., 2016). Thus, there may be a similar relationship for hard-to-detect anomalies.

4.2. Method

4.2.1. Participants

Thirty-two young adults (26 females; M_{age} = 20, age range = 18-26) and 32 older adults (25 females; M_{age} = 70, age range = 65-80) were recruited from University of Leicester and the surrounding area, and course credit or an honorarium was provided for participation.

All participants were Native English speakers, had normal to corrected vision, and did not report any history of dyslexia.

Visual acuity and reading acuity tests were conducted to ensure that older participants did not experience any visual or cognitive impairment beyond what would be expected with normal aging (see Table 5). The Early Treatment Diabetic Retinopathy Study chart (Ferris & Bailey, 1996) was used to assess visual acuity performance. This test was conducted at the recommended 1 metre distance. The Radner reading acuity test was used to assess reading acuity and tested at the recommended 40cm distance (Radner, 2017). Cognitive abilities were examined using the Montreal Cognitive Assessment Test (MoCA) (Nasreddine et al., 2005). Working memory and vocabulary were examined using the Wechsler Adult Intelligence Test (WAIS-IV) (Wechsler, 2008).

Young adults were found to have significantly better visual acuity than older adult participants, $t(62) = 7.44$, $p < .001$, $d = 1.86$ [95% CI 24.10 to 13.90]. Results from the Radner reading acuity test found young adults had better reading acuity than older adults $t(62) = 8.00$, $p < .001$, $d = 2.00$ [95% CI -.12 to -.17]. These results are expected as young adults typically perform better on vision tests compared to older adults (Elliot et al., 1995). There was no significant difference in hours that young and older adults spent reading per week $t(62) = .018$, $p > .05$, $d = 0.00$ [95% CI 3.51 to 3.45], or years spent in education $t(62) = .915$, $p > .05$, $d = 0.23$ [95% CI 1.89 to .70] (see Table 5).

Older adults' working memory scores were marginally higher than young adults' $t(62) = 2.00$, $p = .05$, $d = 0.50$ [95% CI 11.69 to 27.75]. Older adults had better vocabulary scores $t(62) = 4.94$, $p < .001$, $d = 1.24$ [95% CI 9.70 to 4.11]. Cognitive abilities for older adults were examined using the Montreal Cognitive Assessment (Nasreddine et al., 2005) which has a total score of 31. All participants' scores were within the healthy range of ≥ 26 (Means and SD in Table 5).

Table 5

Mean and SD scores of completed education, hours spent reading per week, visual acuity, reading acuity, working memory, vocabulary and MoCA score by age group.

Age	Completed education (years)	Reading (hours per week)	Visual acuity	Radner reading score	Working memory	Vocabulary	MoCA
Young adults	15 (2)	12 (6)	20 (20)	10 (2)	27 (4)	43 (5)	-
Older adults	15 (3)	13 (9)	20 (.13)	14 (2.8)	30 (5)	50 (7)	30 (1)

4.2.2. Design and Materials

The experiment had a 3 (*Anomaly Type*: detected vs. undetected vs. non-anomalous) x 2 (*Age*: young vs. older adults) mixed-design, with Anomaly Type a within-subjects factor and Age as a between-subjects factor.

The experimental items were derived from those used by Bohan (2007). Each experimental passage consisted of three sentences: a pre-target sentence, followed by a target sentence, and finally a post-target sentence (see Example 3). The pre-target sentence provided the context. The target sentence contained a manipulation that rendered the critical word (*victim*) as either anomalous (*prison sentence*) or non-anomalous (*care order*) in relation to the context. The post-target sentence was included so that the target sentence was not at the end of the trial.

3. When you are ready for a drink, this is how to make the perfect cup of tea. (context sentence) \ When using a teabag, pour the (introductory region) \ tea (anomalous manipulation) / water (non-anomalous manipulation)\ carefully and slowly from the (pre-critical region)\ kettle (critical region)\ into a china cup.(post-critical region)\.

3a. Was a china cup used?

The introductory region ranged from four to seven words. The pre-critical region ranged from two to six words. The context manipulation region ranged from one to three words. The character difference in length between anomalous and non-anomalous contexts was zero to four characters. The critical region comprised of one to two words. The post-critical region consisted of three to seven words. Apart from the context manipulation, all regions remained constant across both conditions.

There were two counterbalanced presentation lists, with 18 anomalous and 18 non-anomalous items, resulting in each participant seeing 36 experimental items (See Appendix – Item A). Interspersed with the 36 experimental items were 36 filler items consisting of a variety of syntactic complexities to avoid task effects and prevent readers from adopting strategies (See Appendix – Item B). Thus, participants read 72 items in total. The presentation lists were presented to participants in a pseudorandom order using a Latin square design. All sentences were followed by a yes/no comprehension question (see Example 3a). Overall, half of the questions required a 'yes' answer.

Anomaly detection pre-test: It was crucial to pre-test the materials, especially on an older adult sample as their world knowledge and experience may differ to that of a young sample. In addition, a range of easiness in anomaly detection was required, in order that there would be a sufficient number of 'detected' versus 'undetected' anomalies to enter the eye-tracking analyses. Two norming tasks were conducted as the first task did not generate enough suitable items. Overall, 57 young adults (47 females; M_{age} = 23, age range = 18-30) and 67 older adults took part (41 females; M_{age} = 70, age range = 65-76). Participants read scenarios constructed by Bohan (2007) and were instructed that there would be some "words out of place". Participants were tasked with identifying any "words out of place" and if an anomaly was identified, they were asked to provide a brief explanation. Overall, 86 of the 135 passages used by Bohan (2007) were presented in the pre-test (full list of items in Appendix – Item A). In the first pre-test, 50 items were presented to 40 young and 40 older adults, then in the second pre-test a further 36 items were presented to an additional 17 young adults and 27 older adults. Two lists were created and counterbalanced so that one participant would be presented with one scenario (anomalous or non-anomalous). Item order was randomised in the survey software Qualtrics. Participants completed the pre-test online in their own time, the duration was approximately 20 minutes. Upon completion, participants could opt-in to enter into a prize draw to win a £25 Amazon voucher.

Older adults' performance in detecting semantic anomalies was slightly higher than young adults in the first pre-test ($t(78) = 2.22$, $p < .05$) but no difference was found in the second ($t(78) = 1.02$, $p > .05$). Of the 86 items that were pre-tested, 37 items fell into a central range of 30-85% detection when collapsed by age. As stated above, a range was necessary so that there was enough data to compare detected versus undetected anomalies. A total of 36 items were used for the eye-tracking experiment (full list in Appendix – Item A).

4.2.3. Apparatus

An SR Research Ltd. EyeLink 1000 tower mounted eye-tracker was used to record gaze location every millisecond. The display screen resolution was 1024 x 768 and the refresh rate was 120Hz. Only the right eye was tracked although viewing was binocular. Participants viewed sentences on a 20-inch monitor at a viewing distance of 80 cm with three letters displayed per degree of visual angle. Sentences were displayed in Courier New 14-

point font as black text with two blank lines between each line of text, presented on a white background. A chin and head rest were used to minimise movements.

4.2.4. Procedure

Prior to the presentation of the trials, a nine-point calibration procedure was conducted (average error <.5 degrees of visual angle). At the beginning of the experiment, participants were informed that they should read the items for comprehension after which a yes/no question will be displayed. While the main task was reading for comprehension, the experimenter explained that some passages contained words that were “out of place” and to alert the experimenter if they identified any instances of this. A fixation cross was presented on the left-hand side of the screen at the beginning of each trial. Once participants were correctly fixating on the fixation cross, this was replaced by the initial character of the text. Participants were instructed to press a response key once they had finished reading the item and to press the left button for ‘Yes’ and the right button for ‘No’ when answering the comprehension question. If participants detected an anomaly, they verbally informed the experimenter and described the anomaly identified. The experimenter gave no feedback to any of the answers, simply replying “okay”. After the eye-tracking task, working memory and vocabulary tests were conducted, which was followed by a post-test questionnaire presented in Qualtrics software. This tested the participants’ knowledge of the anomaly, for instance, that a victim is someone who has suffered because of an adverse event (e.g., a crime).

4.2.5. Analyses

The following eye movement measures were analysed: *First-fixation duration* is the duration of the first fixation in each region of interest. *First-pass reading time* is the sum of all fixation durations from the eyes first entering the specific region to first leaving the given region. *Regression path reading time* is the total time from the eyes first entering a region plus all the time a reader fixates to the left of a region before fixating on a subsequent region to the right for the first time. The sum of all fixations in a region is the *total reading time*. *Regressions in* is the proportion of trials in which a regressive saccade is made back into a given region from the right. *Regressions out* describe the proportion of trials in which a regressive saccade is made out of the region during first-pass reading. The *probability of fixating* is the likelihood that the target word is not by-passed on first-pass reading.

Items were categorised as *detected* when the participant correctly identified the anomaly during the eye-tracking task and had accurate knowledge of the anomaly in the post-test questionnaire. Items were categorised as *undetected* when the participant did not correctly identify the anomaly during the eye-tracking task but showed accurate knowledge of the anomaly in the post-test questionnaire. Additionally, in a small percentage of items (7% of anomalies), participants identified the anomaly during the eye-tracking task but then incorrectly answered a question about the item in a post-experiment questionnaire. These items were not included in the analyses as they were not of interest to the aims of the present study. There were no instances where a non-anomalous control was incorrectly detected as anomalous. Finally, *non-anomalous* passages were those which did not contain an anomaly. This was used as a baseline for *detected* and *undetected* anomalies.

The “clean” function in DataViewer (SR Research) was used to trim the data. Fixations shorter than 80ms, and which were located within one character space of the next or previous fixation, were merged into that nearby fixation. As is standard for eye-tracking experiments, fixations shorter than 80ms or longer than 1200ms were discarded. Furthermore, a conservative criterion value of 2.5 standard deviations was set per participant, per condition for each dependent variable and any values above this were removed. In the pre-critical region, 0.32% of first fixation duration times, 0.32% first-pass reading times, 2% regression path reading times and 1.13% total reading times were removed. For the critical region, this procedure accounted for 1.22% of first fixation duration times, 2.56% of first-pass reading times, 1.93% regression path reading times and 1.67% of total reading times. At the post-critical region, 0.7% of first fixation duration times, 0.56% of first-pass reading times, 0.83% regression path reading times and 0.65% total reading times were removed.

4.3. Results

A generalised linear mixed-effects model (glme; Baayen et al., 2008) was employed using the lme4 package (Baayen et al., 2008; Bates et al., 2015) to analyse data using R (Version 3.5.1; R Core Team, 2019). The gamma family and identity link were utilised to reduce the skew of the data (Lo & Andrews, 2015).

All models included Anomaly Type (comprising of detected anomalies, undetected anomalies, and non-anomalous sentences) and Age (young and older adults) which were fixed factors, and items and subjects were random factors. Initially, a full random structure, which included by-subject and by-items slopes with interactions was run to avoid being anti-conservative (Barr et al., 2013). This structure was then trimmed for models that did not converge, with the number of iterations being increased and the optimizer "bobyqa" added. By default, maximum likelihood (ML) estimates were used to compare models. For regressions in, regressions out, probability of skipping, and accuracy, a binomial model was used with the same fixed and random factor procedure described above. Note that the result was constant for the probability of skipping and regressions in to the post-critical region (no skips or regressions in were made), and so, running a GLM analysis was not necessary. Reported results are based on models that successfully converged. *P* values were automatically generated from the lmerTest package (Kuznetsova et al., 2017). Absolute *t*-values (and *z*-values) were considered significant if ≥ 1.96 for all models (see Table 6 for all means and SE). All categorical factors were contrasted using the MASS package (Venables & Ripley, 2002) and function contr.sdiff (0.5/-0.5) meaning that the intercept corresponds to the grand mean and the fixed effects to the main effect of the fixed factors. As Anomaly Type had three levels, a separate model of undetected and non-anomalous sentences was run to obtain a comparison between undetected and non-anomalous sentences. Simple effects were conducted for interactions identified in the GLM (see Appendix – Item C). To assess the relationship between anomaly detection accuracy data and eye movement behaviour, a logistic regression was conducted using the ‘logit’ link function.

Results of detection accuracy are reported first, followed by the relationship between detection accuracy (as an index of shallow processing) and eye movement behaviour (see Table 7 for logit-model results). Then, eye movement data are reported for detected versus undetected comparisons, detected versus non-anomalous comparisons and undetected versus non-anomalous comparisons (see Table 8 and 9 for base model results) (see Appendix – Item D for final model syntax).

Table 6

Means and standard error of young and older adults reading times by age group and region.

Region	Age	Condition	First Fixation Duration	First pass reading times	Regression path reading times	Total reading times	Skipping	Regressions In	Regressions Out
Pre-critical region	Young adults	Undetected	221 (5)	658 (26)	1510 (83)	1701 (70)	0.15 (.02)	0.39 (.03)	0.33 (.02)
		Detected	226 (8)	652 (43)	1312 (88)	2310 (127)	0.17 (.03)	0.63 (.04)	0.34 (.04)
		Non-anomalous	226 (4)	675 (20)	1382 (63)	1646 (52)	0.12 (.01)	0.41 (.02)	0.29 (.02)
	Older adults	Detected	257 (6)	712 (38)	1602 (92)	1683 (72)	0.15 (.02)	0.37 (.03)	0.32 (.03)
		Undetected	241 (6)	630 (27)	1416 (73)	1704 (62)	0.12 (.02)	0.49 (.03)	0.36 (.03)
		Non-anomalous	249 (4)	692 (20)	1310 (47)	1631 (46)	0.15 (.02)	0.40 (.02)	0.30 (.02)
Critical region	Young adults	Undetected	219 (2)	924 (28)	2007 (55)	1875 (46)	0.13 (.01)	0.42 (.01)	0.32 (.01)
		Detected	218 (3)	850 (40)	2235 (103)	2123 (70)	0.16 (.01)	0.55 (.02)	0.33 (.02)
		Non-anomalous	221 (2)	911 (20)	1985 (43)	1866 (34)	0.13 (.01)	0.42 (.01)	0.32 (.01)

	Older adults	Undetected	247 (3)	903 (30)	2098 (66)	1938 (54)	0.14 (.01)	0.42 (.01)	0.32 (.01)
		Detected	241 (2)	833 (26)	1820 (50)	1707 (37)	0.15 (.01)	0.46 (.01)	0.31 (.01)
		Non-anomalous	242 (2)	868 (18)	1899 (40)	1785 (32)	0.15 (.01)	0.42 (.01)	0.32 (.01)
Post-critical region	Young adults	Undetected	226 (7)	645 (28)	4299 (213)	1222 (42)	0.00 (.00)	0.00 (.00)	0.84 (.02)
		Detected	217 (8)	461 (32)	6163 (390)	1519 (83)	0.00 (.00)	0.00 (.00)	0.89 (.03)
		Non-anomalous	228 (4)	642 (22)	4377 (173)	1308 (37)	0.00 (.00)	0.00 (.00)	0.85 (.01)
	Older adults	Undetected	249 (7)	660 (32)	4728 (255)	1417 (63)	0.00 (.00)	0.00 (.00)	0.86 (.02)
		Detected	240 (6)	548 (25)	4060 (181)	1215 (47)	0.00 (.00)	0.00 (.00)	0.84 (.02)
		Non-anomalous	242 (4)	637 (20)	4419 (161)	1307 (38)	0.00 (.00)	0.00 (.00)	0.85 (.02)

4.3.1. Detection accuracy

Older adults were more likely to detect semantic anomalies (M = 54%, SD = 1%) than young adults (M = 28%, SD = 1%), this difference was statistically significant (*Estimate* = 1.948, SE = 0.50, Z = 3.92).

4.3.2. Is there a relationship between eye movement behaviour and shallow processing?

There were several measures of eye movement behaviour that predicted accurate detection of semantic anomalies (see Table 7). At the pre-critical region, longer total reading times were associated with more accurate detection. More regressions into the pre-critical and critical regions also predicted greater accuracy. At the critical and post-critical regions, shorter first-pass reading times were associated with more accurate detection of semantic anomalies. These results suggest that individuals who spent more time revisiting the region of interest were more likely to detect semantic anomalies.

Table 7

Logit-model scores by age group, condition and segment.

Accuracy Measure	Reading measure	β	SE	*t/z*	*p*
Pre-critical region	Intercept	-0.407	0.172	2.37	*
	First fixation duration	0.000	0.000	0.09	
	Intercept	-0.284	0.104	2.75	*
	First pass reading times	0.000	0.000	-1.31	
	Intercept	0.026	0.01	2.83	*
	Regression path reading times	0.000	0.000	1.84	
	Intercept	0.624	0.110	5.70	***
	Total reading times	0.000	0.000	2.60	*
	Intercept	0.673	0.087	7.70	***
	Regressions in	0.613	0.127	4.81	***
	Intercept	0.432	0.078	5.57	***
	Regressions out	0.115	0.132	0.87	
	Intercept	-0.382	0.068	-5.63	***
	Skipping	-0.080	0.181	-0.44	
Critical region	Intercept	-0.429	0.066	6.51	***
	First fixation duration	0.000	0.000	1.04	
	Intercept	-0.032	0.033	9.48	***
	First pass reading times	0.000	0.000	2.50	*
	Intercept	0.038	0.034	9.89	***
	Regression path reading times	0.000	0.000	1.32	

	Intercept	-0.034	0.037	-9.17	***
	Total reading times	-0.000	0.000	-1.07	
	Intercept	-0.500	0.036	-13.66	***
	Regressions in	0.275	0.053	5.17	***
	Intercept	-0.036	0.032	-11.21	***
	Regressions out	0.027	0.057	-0.47	
	Intercept	-0.384	0.029	-13.48	***
	Skipping	0.124	0.075	1.65	
Post-critical region	Intercept	-0.310	0.150	2.06	*
	First fixation duration	0.000	0.000	0.41	
	Intercept	-0.001	0.106	0.06	
	First pass reading times	0.001	0.000	4.26	***
	Intercept	-0.454	0.100	4.57	
	Regression path reading times	0.000	0.000	1.17	
	Intercept	-0.393	0.116	3.38	***
	Total reading times	0.000	0.000	0.284	
	Intercept	0.439	0.168	2.61	**
	Regressions out	0.087	0.182	0.47	
	Intercept				
	Skipping				

*0.001 '***' 0.01 '**' 0.05 '*'*

4.3.3. Detected anomalies versus non-anomalous sentences

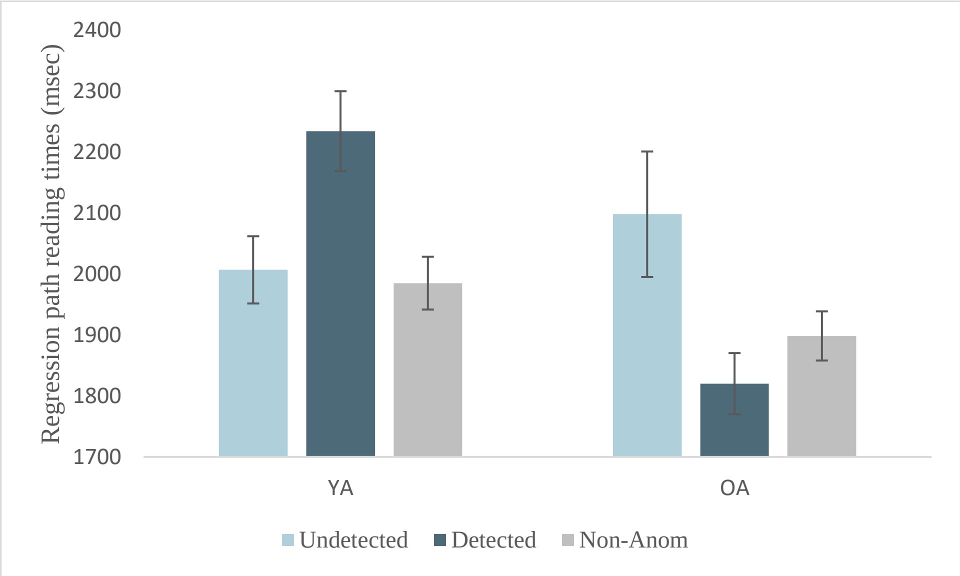
This section describes differences in eye movement behaviour while participants were processing detected anomalies versus non-anomalous sentences. This comparison provides a baseline of differences in controls compared to expected difficulty in processing detected anomalies. In general, the findings showed longer regression path, total reading times, and more regressions when readers were processing detected anomalies compared to non-anomalous sentences. This suggests that as expected, readers experienced greater difficulty when processing detected anomalous compared to non-anomalous sentences. In addition, some age differences were found, and are outlined below.

Older adults had longer first fixation durations compared to young adults at the pre-critical, critical, and post-critical regions. Interactions between Anomaly Type x Age were observed in a number of measures of reading behaviour. Specifically, there was an interaction in first-pass reading times at the critical region that showed both young and older adults had shorter first-pass reading times for detected anomalies than non-anomalous sentences and this effect was larger for young adults in the critical region (older adults = 35ms, young adults = 61ms). This result is somewhat counterintuitive, however, shorter first-pass reading times for difficult conditions can occur when readers immediately make a regression out of the region on encountering the difficulty (thus truncating first-pass reading). At the post-critical region, older adults (548 ms) had longer first-pass reading times than their younger counterparts (461 ms). There were no age differences for non-anomalous control sentences (older adults = 637 ms, young adults = 642 ms).

An interaction was observed between Anomaly Type x Age in regression path reading times at the pre-critical, critical, and post-critical regions. At the pre-critical region participants had not yet encountered the anomaly, nevertheless, older adults (1416 ms) had longer regression path reading times for detected anomalies compared to young adults (1312 ms). Young adults had longer regression path reading times for detected anomalies compared to older adults at the critical (older adults = 1820 ms, young adults = 2235 ms) (see Figure 8) and post critical regions (older adults = 4060 ms, young adults = 6163 ms).

Figure 8

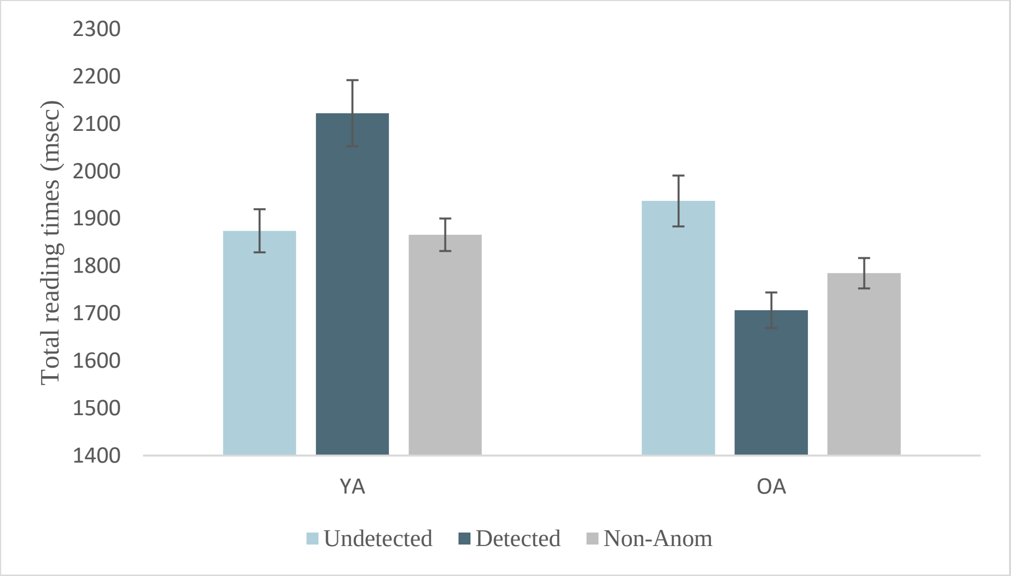
Young and older adults' regression path reading times per condition at the critical region



There was also an Anomaly Type x Age interaction in total reading times at the pre-critical, critical, and post-critical regions. Young adults exhibited longer total reading times for detected anomalies than non-anomalous sentences in the pre-critical region (2310 v 1636 ms), critical region (2123 v 1866 ms) and post-critical region (1519 v 1308 ms). At the pre-critical region, older adults showed no differences between detected and non-anomalous sentence (1704 v 1631 ms). At the critical (see Figure 9) and post-critical regions, an opposite pattern to young adults was observed in that older adults had longer total reading times for non-anomalous sentences (critical region = 1785 ms, post-critical region = 1307 ms) compared to detected anomalies (critical region 1707 ms, post-critical region = 1215 ms).

Figure 9

Young and older adults' total reading times per condition at the critical region

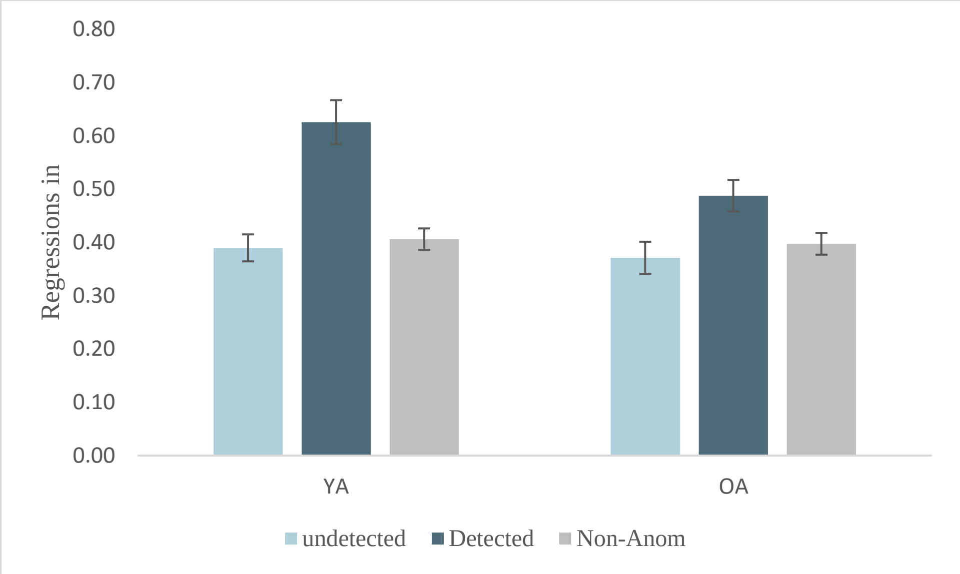


An interaction between Anomaly Type x Age in the probability of skipping at the pre-critical region was found, however no post-hoc tests were significant.

There was an interaction between Anomaly Type x Age in regressions in to the pre-critical and critical regions (see Figure 10), showing that both young and older adults made more regressions in for detected anomalies than non-anomalous sentences. This effect was larger for young adults in the pre-critical (older adults = 0.09, young adults = 0.22) and critical region (young adults = 0.13, older adults = 0.04). This finding supports the other measures indicating that young adults are engaging in re-reading when they detect an anomaly.

Figure 10

Young and older adults' regression into the critical region per condition



4.3.4. Detected versus undetected anomalies

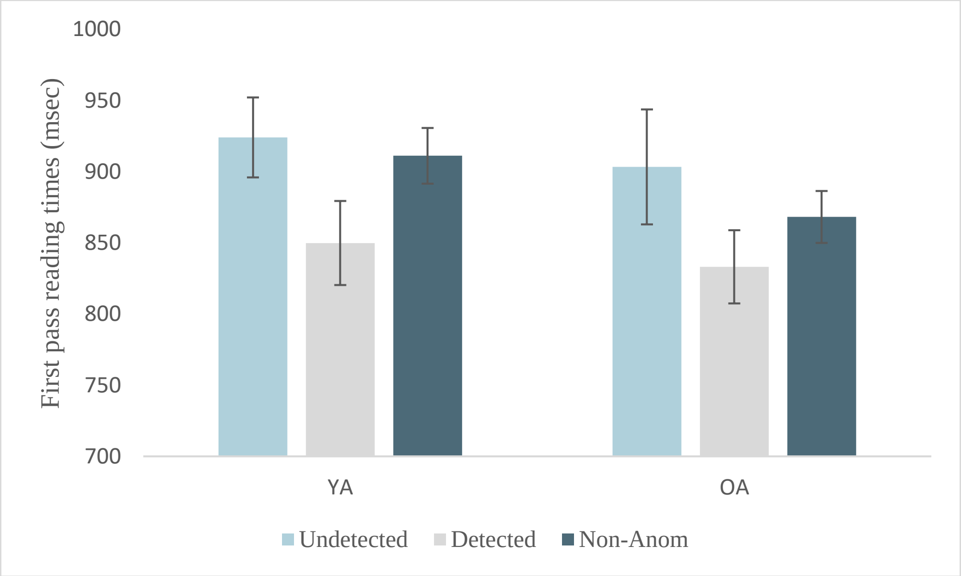
Results identifying differences in eye movement measures between detected versus undetected anomalies are of interest to understand if there are differences in processing between the two conditions. Overall, the results showed that young adults had longer regression path, total reading times, and more regressions were made when processing detected anomalies compared to undetected anomalous sentences. Older adults had longer regression path and total reading times for undetected anomalies suggesting some difficulty was observed but more regressions were made for detected anomalies. These results are described next.

More regressions into the pre-critical region were found for detected anomalies (M = 0.53, SE = .02) than undetected anomalies (M = 0.38, SE = .02), suggesting that participants were engaging in more re-reading for detected than undetected anomalies.

An interaction between Anomaly Type x Age in first-pass reading times at the critical region indicated that both young and older adults had longer first-pass reading times for undetected anomalies (older adults = 903 ms, young adults = 924 ms) than detected anomalies (older adults = 833 ms, young adults = 850 ms) (see Figure 11). At the post-critical region young adults had longer first-pass reading times for undetected anomalies (645 ms) than detected anomalies (461 ms) but there was no difference for older adults (detected = 548 ms, undetected = 660 ms). This eye movement pattern was similar to that observed for detected anomalies versus non-anomalous controls and may illustrate similar behaviour, that is, readers cut short the first-pass reading times to immediately look back to previous information in the more difficult condition.

Figure 11

Young and older adults' first pass reading times per condition at the critical region



An interaction between Anomaly Type x Age was found at the pre-critical, critical, and post-critical regions in regression path reading times. At the pre-critical region, both young and older adults had longer regression path reading times for undetected anomalies than detected anomalies and this effect was larger for young adults (older adults = 198ms, young adults = 186ms). At the critical and post-critical regions, young adults had longer regression path reading times for detected (critical region = 2235 ms, post-critical region = 6163 ms) than undetected anomalies (critical region = 2007 ms, post-critical region = 4299 ms). In contrast, at the critical and post-critical regions, older adults had longer regression path reading times for undetected anomalies (critical region = 2098 ms, post-critical region = 4728 ms) compared to detected anomalies (critical region = 1820 ms, post-critical region = 4060 ms).

An interaction between Anomaly Type x Age was significant for regressions into the critical region. Both young and older adults made more regressions into the critical region for detected than undetected anomalies and this effect was larger for young adults (young adults = 0.13, older adults = 0.04). This finding was similar to the detected versus non-anomalous control results and supports other measures indicating that young adults are re-reading when they detect an anomaly.

Similarly to the detected anomalies versus non-anomalous controls results, an interaction between Anomaly Type x Age for total reading times was identified at the pre-critical, critical, and post-critical regions. Young adults had longer total reading times for detected than undetected anomalies in the pre-critical (detected = 2310 ms v undetected = 1701 ms), critical (detected = 2123 ms v undetected = 1875 ms), and post-critical regions (detected = 1519 ms v undetected = 1222 ms). At the pre-critical region, older adults' total reading times were no different for detected (1704 ms) versus undetected anomalies (1683 ms). At the critical region, older adults showed longer total reading times for undetected (critical = 1938 ms) than detected anomalies (critical = 1707 ms) (see Figure 12). At the post-critical region, there was no difference for older adults (detected = 1519 ms v undetected = 1222 ms).

Figure 12

Young and older adults' total reading times per condition at the post-critical region

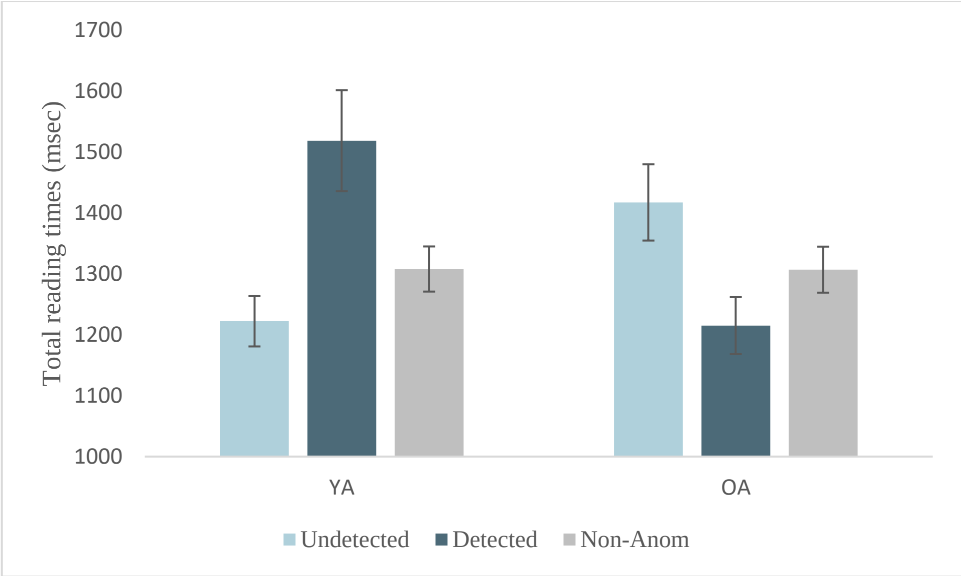

1700
1600
1500
1400
1300
1200
1100
1000
Total reading times (msec)
YA
OA
Undetected
Detected
Non-Anom

Table 8

Linear and binomial mixed-effects models by region and eye-movement measures for detected anomalies compared to undetected anomalies and non-anomalous sentences.

	Pre-critical region				Critical region				Post-critical region			
Measure	Estimate	SE	t/z value		Estimate	SE	t/z value		Estimate	SE	t/z value	
First fixation duration times												
Intercept	239.696	5.655	42.39	***	233.569	4.211	55.46	***	237.509	6.664	35.64	***
Detection v Undetected	0.413	4.922	0.08		1.907	2.053	0.93		-5.124	6.862	-0.75	
Detection v Non-anomalous	0.746	5.377	0.14		-1.155	3.403	-0.34		2.812	7.235	0.39	
Age	25.780	9.666	2.67	**	25.022	5.672	4.41	***	19.158	8.748	2.19	*
Detection v Undetected x Age	-6.100	7.914	-0.77		6.149	14.426	0.43		15.656	9.49	1.65	
Detection v Non-anomalous x Age	1.903	7.500	0.25		-4.049	13.802	-0.29		-13.238	7.908	-1.67	

First-pass reading time												
Intercept	681.228	13.24	51.45	***	874.825	3.75	233.27	***	600.25	14.54	41.29	**
Detection v Undetected	-8.919	11.416	-0.78		-1.657	3.47	-0.48		-93.97	13.21	-7.11	***
Detection v Non-anomalous	12.984	15.112	0.86		12.657	5.11	2.48	*	113.53	12.97	8.75	***
Age	26.045	15.645	1.67		-20.801	2.438	-8.53	***	35.38	11.46	3.09	**
Detection v Undetected x Age	-23.807	14.219	-1.67		87.301	6.789	12.86	***	69.12	17.24	4.01	***
Detection v Non-anomalous x Age	15.767	9.089	1.74		-64.839	4.366	-14.85	***	-69.65	11.22	-6.21	***
Regression path reading times												
Intercept	1422.123	10.051	141.50	***	1992.171	3.333	597.72	***	4695.739	7.075	663.70	***
Detection v Undetected	-20.346	10.855	-1.87		140.518	2.994	46.94	***	997.033	7.347	135.70	***
Detection v Non-anomalous	-73.996	9.231	-8.02	***	-168.828	2.684	-62.91	***	-965.08	7.282	-132.50	***
Age	22.644	12.177	1.86		-137.014	3.486	-39.31	***	-507.157	7.592	-66.80	***
Detection v Undetected x Age	69.829	14.707	4.75	***	-285.312	3.042	-93.78	***	-1533.74	8.905	-172.20	***

Detection v Non-anomalous x Age	-138.819	8.846	-15.69	***	202.731	3.187	63.62	***	1228.165	9.861	124.50	***
Regressions in												
Intercept	-0.224	0.103	-2.19	*	-0.229	0.044	-5.25	***				
Detection v Undetected	0.743	0.176	4.23	***	0.293	0.072	4.06	***				
Detection v Non-anomalous	-0.695	0.200	-3.47	***	-0.320	0.069	-4.66	***				
Age	-0.263	0.160	-1.64		-0.119	0.080	-1.49					
Detection v Undetected x Age	-0.591	0.303	-1.95		-0.304	0.119	-2.56	*				
Detection v Non-anomalous x Age	0.600	0.270	2.22	*	0.315	0.104	3.02	**				
Regressions out												
Intercept	-0.778	0.084	-9.26	***	-0.759	0.034	-22.19	***	1.974	0.131	15.10	***
Detection v Undetected	0.138	0.151	0.92		-0.001	0.062	-0.02		0.184	0.214	0.86	
Detection v Non-anomalous	-0.271	0.165	-1.65		-0.009	0.061	-0.16		-0.157	0.204	-0.77	
Age	0.025	0.131	0.19		-0.026	0.062	-0.41		0.020	0.250	0.08	

Detection v Undetected x Age	0.186	0.295	0.63		-0.060	0.123	-0.49		-0.390	0.425	-0.92	
Detection v Non-anomalous x Age	-0.123	0.264	-0.47		0.075	0.108	0.69		0.348	0.379	0.92	
Total reading times												
Intercept	1827.05	18.48	98.86	***	1873.066	3.364	556.75	***	1397.642	12.232	114.26	***
Detection v Undetected	278.32	13.66	20.38	***	161.98	2.887	56.10	***	259.931	8.891	29.23	***
Detection v Non-anomalous	-353.91	17.95	-19.72	***	-165.612	2.582	-64.15	***	-215.486	8.862	-24.32	***
Age	-191.50	10.88	-17.61	***	-137.021	2.944	-46.54	***	-64.707	21.213	-3.05	**
Detection v Undetected x Age	-401.60	16.29	-24.65	***	-259.455	2.159	-120.17	***	-291.577	13.597	-21.44	***
Detection v Non-anomalous x Age	434.27	19.38	22.41	***	205.951	2.872	71.72	***	231.464	14.165	16.34	***
Skipping Rate												
Intercept	-1.949	0.117	-16.68	***	-1.927	0.081	-23.77	***				
Detection v Undetected	-0.087	0.196	-0.44		0.052	0.119	0.43					

Detection v Non-anomalous	-0.048	0.175	-0.27		-0.060	0.103	-0.58
Age	-0.063	0.226	-0.28		0.038	0.149	0.25
Detection v Undetected x Age	-0.495	0.393	-1.26		-0.204	0.171	-1.19
Detection v Non-anomalous x Age	0.737	0.351	2.10	*	0.273	0.148	1.85

*0 '***' 0.001 '**' 0.01 '*' 0.05 '.'*

4.3.5. Undetected anomalies versus non-anomalous sentences

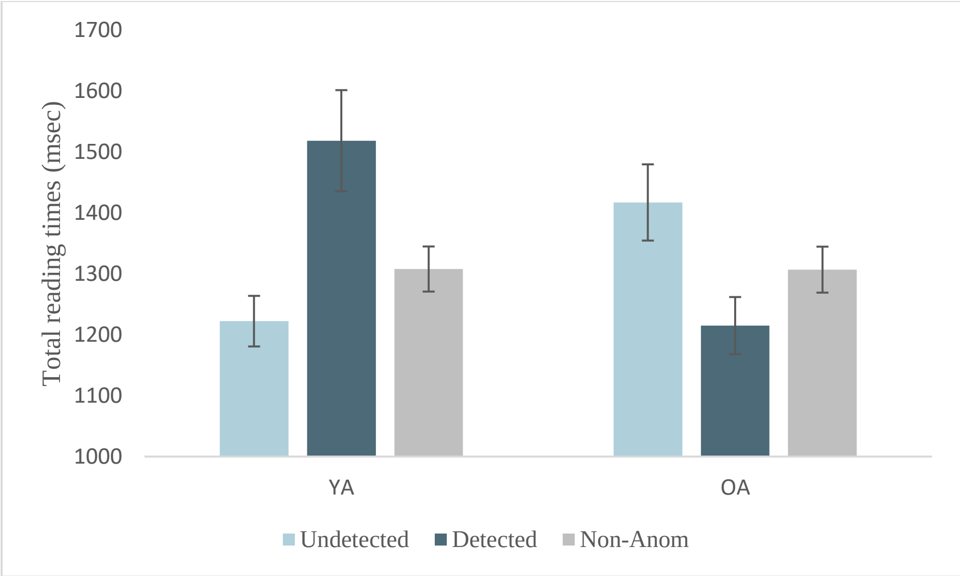
The results for undetected anomalies versus non-anomalous controls are reported to assess whether there is any indication of unconscious processing of undetected anomalies compared to the non-anomalous sentences. That is, if differences are observed, it suggests evidence of unconscious registration. Overall, the results showed some evidence of unconscious registration as longer regression path, and total reading times for undetected anomalies compared to non-anomalous sentences were observed (see Table 9). Age differences are described below.

An interaction between Anomaly Type x Age at the pre-critical, critical, and post-critical regions for regression path reading times showed that older adults had longer regression path reading times for undetected anomalous sentences than non-anomalous controls in pre-critical (undetected = 1602 ms v non-anomalous = 1310 ms), critical (undetected = 2098 ms v non-anomalous = 1899 ms) and post-critical regions (undetected = 4728 ms v non-anomalous = 4419 ms). At all regions, there was no significant difference for young adults between non-anomalous sentences and undetected anomalies.

There was an interaction between Anomaly Type x Age at the critical region for total reading times. Young adults' total reading times were no different in the critical region (non-anomalous = 1866 ms v undetected anomalies = 1875 ms). Older adults had longer total reading times for undetected anomalous sentences (1938 ms) than non-anomalous sentences (1785 ms) in the critical region.

Figure 13

Young and older adults' total reading times per condition at the post-critical region



Total reading times (msec)
1700
1600
1500
1400
1300
1200
1100
1000
YA
OA
Undetected
Detected
Non-Anom

Table 9

Linear and binomial mixed-effects models by region and eye-movement measures for undetected anomalies compared to non-anomalous sentences.

	Pre-critical region				Critical region				Post-critical region			
Measure	Estimate	SE	t/z value		Estimate	SE	t/z value		Estimate	SE	t/z value	
First fixation duration times												
Intercept	239.561	6.419	37.32	***	233.456	2.634	88.64	***	238.846	6.750	35.38	***
Undetected V Non-anomalous	-1.423	4.925	-0.29		-0.196	1.841	-0.11		3.954	6.476	0.61	
Age	26.279	12.842	2.05	*	24.296	2.454	9.90		17.004	8.226	2.07	*
Undetected V Non-anomalous x Age	3.998	8.172	0.49		-1.762	2.011	-0.88		-0.246	7.473	-0.03	
First-pass reading time												
Intercept	681.536	9.361	72.81	***	878.582	2.506	350.64	***	620.755	10.165	61.07	***

Undetected V Non-anomalous	-1.601	10.125	-0.16		-9.879	2.247	-4.40	***	7.839	8.999	0.87	
Age	25.753	8.567	3.01		-31.863	2.250	-14.16	***	11.977	10.289	1.16	
Undetected V Non-anomalous x Age	0.482	9.335	0.05		-7.246	3.260	-2.22	*	-9.004	10.024	-0.90	
Regression path reading times												
Intercept	1425.999	12.764	111.72	***	1957.386	5.322	367.799	***	4479.68	10.54	424.84	***
Undetected v Non-anomalous	97.484	11.311	8.62	***	-15.332	7.374	-2.08	*	-331.78	12.38	-26.8	***
Age	14.377	8.874	1.62		-87.098	2.844	-30.63	***	-186.74	12.63	-14.78	***
Undetected v Non-anomalous x Age	91.92	8.264	11.12	***	63.558	4.23	15.03	***	274.84	13.39	20.52	***
Regressions in												
Intercept	-0.366	0.099	-3.70	***	-0.285	0.044	-6.43	***				
Undetected v Non-anomalous	-0.208	0.171	-1.21		-0.003	0.061	-0.05					
Age	-0.088	0.155	-0.57		-0.038	0.080	-0.48					

Undetected v Non-anomalous x Age	0.004	0.214	0.02		-0.013	0.086	-0.15					
Regressions out												
Intercept	-0.800	0.081	-9.90	***	-0.762	0.033	-23.03	***	1.937	0.126	15.39	***
Undetected v Non-anomalous	0.118	0.150	0.79		0.009	0.054	0.17		-0.024	0.169	-0.14	
Age	0.011	0.129	0.09		-0.017	0.060	-0.28		0.094	0.242	0.39	
Undetected v Non-anomalous x Age	-0.020	0.226	-0.09		-0.026	0.090	-0.29		0.010	0.303	0.03	
Total reading times												
Intercept	1744.889	11.693	149.23	***	1848.914	2.648	698.22	***	1330.903	11.529	115.44	***
Undetected v Non-anomalous	-220.925	14.876	-14.85	***	-50.105	2.429	-20.63	***	-167.674	13.405	-12.51	***
Age	-112.614	8.284	-13.60	***	-92.881	3.266	-28.44	***	-25.293	9.279	-2.73	**
Undetected v Non-anomalous x Age	10.428	12.665	0.823		41.512	2.964	14.00	***	24.47	20.45	1.20	
Skipping Rate												

Intercept	-1.966	0.113	-17.45	***	-1.929	0.079	-24.42	***
Undetected v Non-anomalous	0.103	0.147	0.70		0.017	0.086	0.19	
Age	0.012	0.221	0.06		0.074	0.147	0.50	
Undetected v Non-anomalous x Age	-0.341	0.293	-1.16		-0.102	0.124	-0.82	

*0 '***' 0.001 '**' 0.01 '*' 0.05 '.'*

4.4. Discussion

In the present study, we recorded eye movements during reading to investigate age differences in the online processing of sentences containing hard-to-detect semantic anomalies. There was a particular focus on whether young and older adults' eye movements showed evidence of unconscious registration of undetected anomalies, despite anomalies going unreported. Additionally the relationship between the eye movements during online processing and semantic anomaly detection was analysed. Young and older adults' working memory and vocabulary was assessed using the WAIS-IV but did not converge in R or was not deemed a better fit than the base model.

The shallow processing account suggests that undetected anomalies will not be processed any differently to non-anomalous sentences (Bohan et al., 2012; Bohan & Sanford, 2008). Furthermore, older adults were expected to engage in more shallow processing when reading anomalous sentences compared to their younger counterparts as recall and integration of information may become increasingly difficult due to cognitive decline during aging (Salthouse, 1990) and so older adults may opt to rely on the greater experience and knowledge that is accumulated across the lifespan.

The results demonstrated clear and consistent differences in reading behaviour on sentences containing detected anomalies compared to non-anomalous and undetected anomalous sentences - indicating that detected semantic anomalies resulted in more disruption during online processing in all regions of analysis. There was some evidence of unconscious processing of undetected anomalies as undetected anomalies produced longer reading times in later eye-movement measures compared to non-anomalous sentences. Furthermore, several age differences were found. The eye movements of young adults suggested that this group had greater difficulty in processing semantic anomalies, often having longer reading times (especially regression path and total reading times) and making more regressions back to earlier parts of the text for detected anomalies compared to non-anomalous controls. Older adults' eye movements during reading indicated more unconscious registration of undetected anomalies versus non-anomalous sentences than young adults. These results are explored further in the following sections.

4.4.1. Eye movement data

In this study, we explored whether any age differences occur in the online processing of sentences containing hard-to-detect semantic anomalies. Daneman et al. (2006) suggested older adults were more efficient in semantic anomaly detection as older adults identified the anomaly before young adults and then made more regressions back to previous regions to resolve the anomaly compared to young adults. In contrast, the results of the current study showed that both young and older adults were inclined to make regressions back to earlier parts of the sentence to immediately resolve the anomaly, and if anything, these effects were larger for young adults.

Differences between undetected anomalies and non-anomalous controls were examined and were particularly important to understand if undetected anomalies are unconsciously registered despite not being reported. Previous studies have examined unconscious processing of anomaly detection and found mixed results, with some studies finding no evidence of unconscious processing of undetected anomalies (Bohan & Sanford, 2008; Sanford et al., 2011) and some studies finding evidence of unconscious processing (Cook et al., 2016; Hannon, 2015; Reder & Kusbit, 1991; Van Oostendorp & De Mul, 1990). The current findings indicate evidence of some unconscious processing, especially in older adult readers. Older adults had longer total reading times in the critical region and longer regression path reading times in the critical and post-critical regions for undetected anomalies compared to non-anomalous controls. Furthermore, older adults showed evidence of processing detected and undetected anomalies differently with, surprisingly, longer reading times found for undetected anomalies compared to detected anomalies. There was little evidence of unconscious processing for young adults as comparisons showed young adults exhibited longer reading times for detected anomalies compared to undetected anomalies and non-anomalous controls than undetected anomalies.

In relation to frameworks that explain reading comprehension, the shallow processing framework (Sanford & Sturt, 2002), the mechanisms proposed by Hannon and Daneman (2001), and the RI-Val model (Cook et al., 2016) share similar underlying explanations regarding the detection of semantic anomalies. The shallow processing framework and the mechanisms described by Hannon and Daneman (2001) suggest that the first criteria in detection is the retrieval of lexical information or knowledge, followed by the integration of information (Daneman et al., 2006b; Sanford & Sturt, 2002). In contrast, the RI-Val model proposes three stages; resonance, integration, and validation (Cook et al., 2016). Similarly to other accounts, new text is linked to context within the sentences and one’s world knowledge, which is then integrated with information in working memory, this is then validated. This validation stage is a low-level pattern matching process (Cook et al., 2016). All three stages are believed to be passive and asynchronous, and unlike other accounts, the RI-Val model includes a coherence threshold. In this threshold, readers may use the information gathered to reach a threshold that is *good enough* and allows them to move their attention to other areas of the sentence. The RI-Val model suggests that undetected sentences are expected to be processed online in later stages and may go undetected because the differentiating detail is not successfully processed either at the resonance, integration or validation stage (Cook et al., 2016).

The findings in the current study highlight that the failure to detect anomalies may be due to integration (Cook et al., 2016; Daneman et al., 2006b; Sanford & Sturt, 2002) or validation (Cook et al., 2016). Our findings show that young and older adults may engage in different eye movement behaviour during semantic anomaly detection. In the present study, young adults were more likely than older adults to validate information producing more regressions into the pre-critical region for detected anomalies compared to non-anomalous and undetected anomalies. Older adults' eye movements suggested less difficulty in the detection of semantic anomalies compared to young adults as older adults showed longer reading times for undetected anomalies compared to detected anomalies. Instead, older adults spent more time engaged in reprocessing as this group had longer regression path reading times in all regions and longer total reading times in the post-critical region for undetected anomalies suggesting more time was spent integrating information, however integration was not successful. Older adults' linguistic experience may act as a compensatory factor enabling them to better validate information, resulting in greater detection of semantic anomalies compared to young adults. Generally, the results support the RI-Val model, which suggests that anomalies go undetected when incomplete integration or retrieval with the wider discourse occurs (Cook et al., 2016).

4.4.2. Is there a relationship between eye movement behaviour and shallow processing?

Examining the relationship between eye movements and comprehension accuracy is crucially important in order to gain a fuller understanding of the processes underlying language comprehension (Ferreira et al., 2019). We examined whether there is a relationship between depth of processing and online reading behaviour, specifically, detection accuracy and eye movement measures were inputted into a logistic regression. At the pre-critical region, longer total reading times and more regressions in predicted greater accuracy, as did more regressions into the critical region. At the critical and post-critical regions, shorter first-pass reading times predicted more accurate detection. It may be the case that shorter first-pass reading times in the post-critical region are associated with fewer errors as these are the trials where people are looking back more (and thus truncating their first-pass reading times on this region).

These findings are compatible with previous research which has found that more rereading and regressions may be associated with successful comprehension (Schotter et al., 2014; Weiss et al., 2018). Previously, one study has examined the relationship between shallow processing and eye movement behaviour. Wonnacott et al. (2016) measured the relationship between eye movement behaviour and whether this predicts shallow processing at the sentence-level using temporary syntactic ambiguities. In this study, children and young adults' eye movements were recorded. The number of regressions made into earlier regions of text positively predicted comprehension, however, note that this was the case for both ambiguous and unambiguous conditions. Interestingly, Wonnacott et al. (2016) found no relationship between total reading times and shallow processing, however, in the present study total reading times at the pre-critical region were linked to successful semantic anomaly detection. This additional time may be indicative of more time spent re-processing which may lead to successful anomaly detection. The results from this study contribute to the understanding of the relationship between online processing and offline comprehension showing that in addition to regressive eye movement behaviour, total reading times are also a predictor of successful comprehension.

4.4.3. Conclusion

Experiment 1 investigated the processing of hard-to-detect anomalies in young and older adults. The findings demonstrated evidence of unconscious registration of semantic anomalies occurring despite the anomaly not being reported, particularly in older adults. These findings are consistent with the RI-Val model of reading comprehension (Cook et al., 2016; Hannon, 2015). A second aim was to examine the relationship between online processing and offline comprehension. The results from the current study suggest that individuals who are more likely to look back and spend more time re-processing earlier regions of text are more likely to correctly detect semantic anomalies, thus furthering our knowledge of the relationship between moment-to-moment processing of text and successful comprehension.

In the next chapter, Experiment 2 extends the current findings by exploring the age differences in good enough processing at the sentence level. Eye movements were recorded to investigate the online processing of sentences containing temporary syntactic ambiguities in young and older adult readers. The distance between the ambiguous portion of text and the subsequent disambiguating information was manipulated to understand if the longer a reader holds on to the initial misinterpretation affects ultimate comprehension. By conducting this experiment on young and older adults, I was able to explore the age differences in whether older adults are more likely to hold on to lingering misinterpretations. A further aim of this experiment was to establish if there is a relationship between eye movements during reading and comprehension accuracy. Similar to previous experiments, working memory, and vocabulary abilities were recorded for all participants.

Chapter 5: Temporary syntactic ambiguities

5.1. Introduction

Standard models of reading comprehension assume that syntactic structure is built incrementally, independently of pragmatic information, to produce a complete representation of the text (see, e.g., Chomsky, 1965; Ferreira & Clifton, 1986; Frazier & Rayner, 1982; MacDonald et al., 1994; Tanenhaus et al., 1995). However, by comparison with this approach, shallow processing (Sanford & Sturt, 2002) and 'good enough' processing accounts (Ferreira et al., 2002) advocate that readers do not always conduct a complete syntactic analysis, instead they rely on ‘fast and frugal’ heuristics which result in a partially complete representation that is good enough for the task at hand.

Good enough processing has been most widely investigated using sentences containing temporary syntactic ambiguities, such as (1a).

1a. While the gentleman was eating the burgers were still being reheated in the microwave. *(temporarily ambiguous)*

1b. While the gentleman was eating, the burgers were still being reheated in the microwave. *(unambiguous control)*

Question: Was the gentleman eating the burgers? Yes/No

Sentences such as (1a) can be temporarily misleading because when readers encounter the phrase *While the gentleman was eating the burgers*, the phrase *the burgers* initially is processed as the direct object of *eating* (Jacob & Felser, 2016; Pickering & Traxler, 1998). Then, when the parser encounters structural information that is inconsistent with this analysis (at *were*), this phrase should be reprocessed as the subject of the main clause (i.e., *in the* *microwave*). Furthermore, when encountering the second verb (*reheated*), a semantic inconsistency with the direct-object analysis is also generated - the gentleman cannot be both eating the burgers and the burgers being reheated - which should further support reanalysis. By comparison with the example in (1a), the sentence in (1b) is structurally unambiguous because the comma after *eating* indicates that *the burgers* is not the object but the subject of the sentence.

Comprehension questions presented after these items can be used to measure the readers' derived interpretation. For example, after reading examples (1a) and (1b), readers might be asked: "*Was the gentleman eating the burgers?*" If a reader answers "no" to this question, this might imply that they have interpreted the sentence correctly. However, if they answer "yes", it suggests the presence of a lingering misinterpretation (Christianson et al., 2001; Jacob & Felser, 2016; Slattery et al., 2013; Sturt, 2007) in which an incorrect initial misanalysis of the ambiguity (in which the gentleman was eating the burgers) is not rejected despite later information in the sentence disambiguating the ambiguity in favour of an alternative interpretation. In addition, measures of eye movements can provide us with insights into how this information is processed during reading. Typically, when reading temporarily ambiguous sentences like (1a) compared to the same phrase in unambiguous sentences like (1b), young adults have longer reading times for syntactically disambiguating information (e.g., the phrase *were still*) and are more likely to reread the disambiguation phrase and make more regressions (i.e., backwards eye movements) back to this region (Jacob & Felser, 2016).

In the previous chapter, age differences in good enough processing at the word level were examined and older adults were less likely to engage in good enough processing than their younger counterparts. The focus of this chapter is to examine whether age differences occur in good enough processing at the sentence level. There are two reasons why older adults may be of interest. Firstly, older readers have been found to experience greater difficulty in processing text (Rayner et al., 2006), especially when text is syntactically complex (Obler, 1991). Secondly, research has shown that world knowledge typically increases across the lifespan (Schaie, 1994) and older adults often perform better on vocabulary tests than their younger counterparts (Ben-David et al., 2015; Keuleers et al., 2015). Coupled together, older adults may compensate for any difficulty in processing text by utilising their world and linguistic knowledge to make inferences based on everyday language use.

If age differences occur, this may be due to a decline in working memory capacity, which occurs naturally during aging (Salthouse, 2012; Salthouse & Babcock, 1991). One way to examine the effect of working memory in good enough processing is to increase the distance between the temporarily ambiguous phrase and the subsequent disambiguating information, for instance, by including additional text, such as *that were really huge* in 1c and 1d compared to 1a and 1b:

1a. While the gentleman was eating the burgers were still being reheated in the microwave. *(short, temporarily ambiguous)*

1b. While the gentleman was eating, the burgers were still being reheated in the microwave. *(short, unambiguous control)*

1c. While the gentleman was eating the burgers that were really huge were still being reheated in the microwave. *(long, temporarily ambiguous)*

1d. While the gentleman was eating, the burgers that were really huge were still being reheated in the microwave. *(long, unambiguous control)*

Question: Was the gentleman eating the burgers? Yes/No

It has been argued that the longer a parser holds the initial misinterpretation (i.e., if there is a longer distance between the ambiguous portion of text and subsequent disambiguating information), the less likely it is that reanalysis will occur, possibly due to working memory constraints (Christianson et al., 2001; Ferreira & Henderson, 1991). Christianson et al. (2006) conducted a self-paced reading study where sentences like (1a – 1d) were presented to examine age differences in processing sentences containing temporary syntactic ambiguities. In this example, 1c and 1d contain an extra clause which creates distance between the ambiguous portion of text and subsequent disambiguating information. The results from this experiment showed no age differences in reading times or accuracy when the length of sentences was manipulated. Working memory capacity was also assessed using Daneman and Carpenter's (1985) reading span task. No correlation between reading span and accuracy in young adult readers was identified (Experiment 3), whereas with older adult readers, the higher the older adults' reading span the greater their accuracy scores. In contrast to Christianson et al.'s (2006) study, other self-paced reading task studies, such as DeDe's (2014) study found working memory was not a significant predictor of aging effects. One important limitation of studies using a self-paced reading task is that this method does not allow participants to read naturally. In contrast eye-tracking records the moment-to-moment eye movements of readers whilst they read sentences naturally and so this was the method used in the current experiment.

There are few studies that have measured the eye movements of young and older adults when reading temporary syntactic ambiguities (Kemper et al., 2004; Kemtes & Kemper, 1997; Payne & Stine-Morrow, 2014). Kemper et al. (2004) conducted two eye-tracking experiments investigating age differences and working memory span differences in processing temporarily ambiguous (e.g., *Several angry workers warned about low wages decided to file complaints*) and unambiguous (e.g., *Several angry workers who were told* *about low wages decided to file complaints*) sentences. In Experiment 1, age differences were assessed, and results at the disambiguating region showed that both young and older adults exhibited longer reading times for garden-path sentences compared to control sentences, suggesting both age groups were similarly garden-pathed initially. Additionally, older adults made more regressions back to the subject noun phrase (e.g., *Several angry workers*) for ambiguous sentences than unambiguous sentences, whereas younger adults did not. In Experiment 2, high and low span younger adults and high and low span older adults were presented with the same sentences as in Experiment 1. Working memory span was derived from the WAIS-R (Wechsler, 1958) and the Daneman and Carpenter reading span task (Daneman & Carpenter, 1980). No age differences were found when memory span was matched. This section highlights that studies that have investigated the role of working memory between young and older adults when reading temporary syntactic ambiguities have found mixed results. In the current study, our design includes a linguistic assessment of working memory by creating distance between the ambiguous portion of text and subsequent disambiguating information (like 1c and 1d) and assesses working memory directly using one of the most widely used assessments, the WAIS-IV (Coalson et al., 2010) in order to better understand whether working memory mediates age differences in good enough processing.

5.1.1. Present study and predictions

The aim of the present study is to examine younger and older adults' eye movements while they read sentences containing a temporary syntactic ambiguity (such as 1a-d). Additionally, I will investigate the influence that the distance between the ambiguous information and the syntactic disambiguation phrase has on good enough processing, as working memory may modulate good enough processing, especially in older adult readers.

The good enough processing framework postulates that a reader will utilise fast and frugal heuristics, this is often 'good enough' for the task at hand and readers want to maintain a state of cognitive equilibrium. In relation to performance on comprehension questions, following Christianson et al.'s (2006) findings, it was predicted that as a result of lingering misinterpretations not being successfully resolved, a lower proportion of correct responses to comprehension questions for ambiguous sentences compared to unambiguous sentences would be found for both young and older adults. Whilst one would expect both groups to engage in good enough processing, young adults are anticipated to be more likely to reanalyse the syntactic and semantic disambiguation phrases and so make less errors on comprehension questions than older adult readers. Older adults are expected to have a greater reliance on world knowledge, coupled with the greater difficulty that this group have been found to experience when reading sentences, older adults are predicted to make more errors on comprehension questions.

In terms of eye movement predictions, it is expected that longer reading times will be observed for ambiguous compared to unambiguous sentences for both younger and older adults, meaning initially both groups will be similarly garden-pathed (Kemper et al., 2004). Age differences are anticipated to occur in reanalysis of the syntactic information, young adults may make a greater number of regressions out of the syntactic disambiguation region compared to older adults and spend more time rereading.

If working memory mediates age differences in good enough processing, there may be a greater ambiguity effect in "short" than "long" sentences at the syntactic disambiguation region. Young and older adults with low working memory capacity may have shorter reading times and make fewer regressions in to or out of the disambiguation regions compared to young and older adults with high working memory scores. Furthermore, if readers are more likely to hold on to their initial incorrect interpretations for "long" than "short" sentences, this should be reflected in a lower proportion of correct responses to comprehension questions following "long" compared to "short" sentences (Ferreira & Henderson, 1991). Again, it is expected that this pattern will be especially prevalent in older adult's accuracy scores.

5.2. Method

5.2.1. Participants

Ninety-six native English-speakers participated: 48 young adults (39 females; M_{age} = 20, age range = 18-26) and 48 older adults (30 females; M_{age} = 70, age range = 65-77). Participants were recruited from the University of Nottingham and surrounding area and received course credit or an honorarium for taking part in the experiment. All participants had normal or corrected vision and did not report any history of dyslexia. The study was approved by the University of Nottingham's School of Psychology ethics board (Ref: S1032R) and all participants gave written consent of their participation.

A variety of tests were conducted to check that older adult participants did not have visual or cognitive impairment beyond what is typical in normal aging. Specifically, visual acuity was assessed using a SLOAN ETDRS chart, and contrast sensitivity was assessed using a SLOAN eye chart. Each of these tests were conducted at the recommended distance of 40cm. Working memory (WM) was examined using the Wechsler Adult Intelligence Test (WAIS-IV) (Wechsler, 2008). Cognitive abilities in older adult participants were examined using the Montreal Cognitive Assessment (MoCA) Test (Nasreddine et al., 2005) (see Table 10).

Young adults had better visual acuity, $t(94) = 6.69$, $p < .001$, $d = 1.97$ [95% CI -9.50 to -5.15], and contrast sensitivity, $t(94) = 7.17$, $p < .001$, $d = 3.22$ [95% CI -19.87 to -11.25], than older adults. Older adults typically perform worse on vision tests (Elliot et al., 1995). Importantly, all participants had visual acuity of up to 20/40, which is within clinically acceptable limits, and all older adults scored within the healthy range on the MoCA (≥ 26) (see Table 10).

5.2.2. Design and Materials

A 2 (*Length*; short vs. long) x 2 (*Ambiguity*; ambiguous vs. unambiguous) x 2 (*Age*; younger vs. older adults) mixed design was employed, with Length and Ambiguity as within-subjects factors, and Age as a between-subject factor. There were 32 experimental items. Twenty were derived from those used by Jacob and Felser (2016) and an additional 12 were constructed in the same style (see Table 11 for an example, and the Appendix – Item E for the full list of stimuli). Each experimental item consisted of three sentences; a context sentence, followed by the target sentence, and finally a post-target sentence. The target sentences for "short" temporarily ambiguous items and their unambiguous controls (e.g., 11a and 11b in Table 11) both started with a noun phrase (e.g., *While the child was painting*). Unambiguous controls included a comma after the subordinate verb (e.g., *painting,*), which facilitated the subsequent noun phrase (e.g., *a truck*) being processed as the subject. An auxiliary (e.g., *was*) followed the noun phrase. This word, along with the following adverb, acted as the syntactic disambiguation phrase (e.g., *was slowly*). The next auxiliary (e.g., *being*) and past participle (e.g., *loaded*) followed the syntactic disambiguation phrase and created a semantic disambiguation phrase. Finally, a prepositional phrase (e.g., *in a garage nearby*) was included to measure sentence wrap-up effects. The "long" items (e.g., 11c and 11d in Table 11) contained an additional relative clause (e.g., *that was quite large*) between the noun phrase and syntactic disambiguation phrase. Experimental items were followed by binary yes/no comprehension questions (e.g., *Was the child painting a truck?*). This offline measure was used to assess whether readers had engaged in good enough processing (i.e., had a lingering misinterpretation). If participants answered 'no', it indicated that successful reanalysis had occurred and if readers answered ‘yes’, this was considered evidence of good enough processing.

Table 10

Means (M) and standard deviations (SD) of visual acuity and working memory measures by age group.

Age	Education (years)	Reading (hr/week)	Visual acuity (Snellen)	Contrast sensitivity (Snellen)	Working Memory (WAIS-IV)	MoCA
Young adults	14 (1)	12 (10)	20 (4)	26 (5)	10 (2.81)	
Older adults	16 (3)	11 (8)	27 (6)	41 (14)	12 (2.01)	30 (1)

Note: MoCA = Montreal Cognitive Assessment (MoCA). Visual acuity: R = 40 – 12, contrast sensitivity: R = 60 – 16, MoCA: R = 31-26. Clinical range for MoCA is ≥ 26. The maximum score was 31.

Table 11

Example material in all conditions, with analysis regions included.

	Condition: Ambiguity	Condition: Length	Example Material
11a.	Ambiguous	Short	The mother entered the room (context sentence). While the child was painting a truck (noun phrase) was slowly (syntactic disambiguation phrase) being loaded (semantic disambiguation phrase) in a garage nearby (wrap-up phrase). The mother told her son that dinner was ready (post-target sentence).
11b.	Unambiguous	Short	The mother entered the room (context sentence). While the child was painting, a truck (noun phrase) was slowly (syntactic disambiguation phrase) being loaded (semantic disambiguation phrase) in a garage nearby (wrap-up phrase). The mother told her son that dinner was ready (post-target sentence).
11c.	Ambiguous	Long	The mother entered the room (context sentence). While the child was painting a truck that was quite large (noun phrase) was slowly (syntactic disambiguation phrase) being loaded (semantic disambiguation phrase) in a garage nearby (wrap-up phrase). The mother told her son that dinner was ready (post-target sentence).

11d.	Unambiguous	Long	The mother entered the room (context sentence). While the child was painting, a truck that was quite large (noun phrase) was slowly (syntactic disambiguation phrase) being loaded (semantic disambiguation phrase) in a garage nearby (wrap-up phrase). The mother told her son that dinner was ready (post-target sentence).

In addition to the 32 experimental items, participants read 64 ‘filler’ items to help prevent unnatural reading strategies as a result of task effects (see Appendix – Item F). Twenty-four of these filler items were the materials for (Chapter 7). The remaining 40 fillers consisted of a variety of different syntactic constructions, of which 15 were ‘pseudo-experimental’ items (see Appendix – Item F). These items began in a similar format to the experimental materials but did not include a syntactic or semantic ambiguity. In addition to questions presented after the 32 experimental items, some of the filler items were also followed by questions, resulting in half of all items being followed by a question. Across the whole set of items seen by participants, an equal number of ‘yes’ and ‘no’ responses were required. There were four counterbalanced presentation lists resulting in participants seeing eight items in each of the four experimental conditions, along with the 64 fillers. Materials were displayed in Courier New 14-point font as black text on a white background and were presented in a different random order for each participant.

5.2.3. Procedure

An SR Research EyeLink 1000 desktop mounted eye-tracker was used to record gaze location every millisecond. Only the right eye was monitored although viewing was binocular. A chin and forehead rest was used to minimise movements. Participants viewed sentences on a 17-inch CRT monitor at a viewing distance of 58cm. Three characters subtended approximately one degree of visual angle. At the beginning of the experiment, participants were informed that they should read the sentences for comprehension, after which a yes/no question may be displayed. Prior to the presentation of the trials, a nine-point calibration procedure was conducted (average error <.5 degrees of visual angle), and the accuracy of the calibration was checked before the presentation of each item. At the beginning of each trial, participants fixated on a box located on the left-hand side of the screen, after which the stimulus was automatically presented and the box was substituted by the first letter of the item. Each item was presented across three lines, with double line spacing between each line of text. Instructions presented to all participants comprised of selecting a button which was situated on a PC game pad to indicate that they had finished reading. The item then disappeared and was replaced by a question, or if there was no question, a calibration target. To answer the questions, participants pressed the left button for ‘no’ and right button for ‘yes’ on the PC game pad. Between each trial, a calibration target appeared. A break was offered every third of the way through the trials and it took approximately 45 minutes for younger adults to complete and one hour for older adults.

Prior to the eye-tracking session, participants completed the visual tests, cognitive assessments, and demographic questions. After the eye-tracking session, participants completed the WAIS-IV digit-span forward, digit-span backward, and digit-span sequential tests for a comprehensive working memory score. The whole experiment took younger adults one and a half hours to complete and older adults approximately two hours.

5.2.4. Analyses

Prior to analysis, short contiguous fixations were combined in an automatic procedure. Specifically, fixations under 80 ms were integrated into larger adjacent fixations within one character, and fixations under 40 ms that were not within three characters of another fixation were deleted, as were fixations over 1,200 ms. Trials where participants failed to read the sentence or where track-loss occurred were eliminated. Specifically, trials where two or more adjacent regions had zero first-pass reading times were removed, accounting for 1.62% of the data.

A number of measures of eye movement behaviour are reported for the syntactic disambiguation (e.g., *was slowly*), semantic disambiguation (e.g., *being loaded*), and sentence wrap-up (e.g., *in a garage nearby*) regions. Measures of initial processing difficulty were *first fixation duration*, which constitutes the duration of the first fixation in each region of interest, and *first-pass reading time,* which is the sum of all fixation durations from the eyes first entering the region to first leaving it. *Regression path reading time* is the sum of fixation durations from the eyes first entering a region to the eyes fixating on a subsequent region for the first time. Regression path reading times are indicative of time spent resolving issues encountered and integrating information with the context. The sum of all fixations in a region is the *total reading time* measure, which provides an indication of overall processing difficulty. *Regressions out* reflects the proportion of trials in which a regressive saccade is made out of the region during first-pass reading. *Regressions in* is the proportion of trials in which regressive saccades are made back into the region and provides information regarding whether text was revisited. The *probability of fixating* was used as an index of skipping and indicates the proportion of trials in which a region is fixated during first-pass reading.

Prior to analysis, all trials with zeros for each continuous reading time measure in each region were removed from that particular analysis. Furthermore, a conservative criterion value of 2.5 standard deviations was set per participant, per condition for each dependent variable and any values above this were removed. This procedure accounted for 8.91% of first-pass, 9.11% of regression path, and 3.90% of total reading time data in the syntactic disambiguation region. In the semantic disambiguation region, 0.71% of first-pass, 0.84% of regression path, and 0.34% of total reading times were removed. Finally, at the sentence wrap-up region, this procedure accounted for a loss of 3.02% of first-pass, 0.84% of regression path, and 0.98% of total reading times.

5.3. Results

A generalised linear mixed effects model (glme; Baayen et al., 2008) was conducted using the lme4 package (Baayen et al., 2008; Bates et al., 2015) to analyse the data in R (Version 3.5.1; R Core Team, 2019). The gamma family and identity link were utilised to reduce the skew of the data (Lo & Andrews, 2015). All categorical factors were contrasted using contr.sdiff.

All base models included Ambiguity, Length, and Age as fixed factors and subjects and items were random factors. Initially, all models included a full random structure by-subject and by-items slopes to avoid being anti-conservative. If a model failed to converge, the number of iterations was increased and the optimizer “bobyqa” was used followed by trimming of models until convergence was successful (Barr et al., 2013). *P* values were automatically calculated from the lmerTest package (Kuznetsova et al., 2017). For regressions in, regressions out, probability of skipping, and comprehension accuracy, a binomial model was used with the same fixed and random factors as above. Absolute *t*-values (and *z*-values) equal to or greater than 1.96 were considered significant. All categorical factors were contrasted using contr.sdiff using the MASS package (Venables & Ripley, 2002), this meant that the intercept was in accordance with the grand mean and the fixed effects matched the main effects of the fixed factors. Simple effects were conducted for interactions identified in the GLM (see Appendix – Item G).

By default, maximum likelihood (ML) estimates were used to compare models. Separate models were fitted for the covariate, Working Memory (WM). As this was a continuous variable, the factor was centered prior to running the GLM. Models with the WM covariate did not converge. Only models with successful convergence are reported (see Appendix – Item D for model syntax).

5.3.1. Comprehension accuracy

The accuracy of responses to comprehension questions was analysed to determine whether young or older adults were more likely to engage in good enough processing. If participants engaged in good enough processing, a higher proportion of incorrect 'yes' answers would be made to comprehension questions (see Table 12 for descriptive statistics). Overall, results showed that good enough processing was common (*M* = 53% correct, *SE* = 1%) and the proportion of correct answers for experimental items was not significantly different between young (*M* = 55%, SE = 1%) and older adults (*M* = 50%, *SE* = 1%).

There was a significant main effect of length that showed "long" sentences (*M* = 49%, *SE* = 1%) resulted in a lower proportion of correct answers than "short" sentences (*M* = 56%, *SE* = 1%) (see Table 13). A main effect of ambiguity showed that ambiguous sentences (*M* = 50%, *SE* = 1%) resulted in a lower proportion of correct answers than unambiguous sentences (*M* = 55%, *SE* = 1%). These main effects were qualified by a significant Length x Ambiguity interaction (see Table 13). The interaction showed that whereas "long" ambiguous sentences (*M* = 45%, *SE* = 2%) resulted in fewer correct responses than "short" ambiguous sentences (*M* = 56%, *SE* = 2%), there were no significant length effects for unambiguous sentences (long ambiguous sentences: *M* = 54%, *SE* = 2%; short ambiguous sentences: *M* = 56%, *SE* = 2%). This result suggests that the distance between the ambiguous noun phrase and the syntactic disambiguation region resulted in greater difficulty in generating the correct interpretation, but only when sentences were ambiguous.

Table 12

Means and standard error for accuracy by condition and age group.

Age	Ambiguity	Length	Accuracy
Younger adults	Ambiguous	Long	47 (3)
		Short	58 (3)
	Unambiguous	Long	56 (3)
		Short	59 (3)
Older adults	Ambiguous	Long	43 (3)
		Short	53 (3)
	Unambiguous	Long	52 (3)
		Short	54 (3)

Table 13

Binomial mixed effect of accuracy, condition and age group.

Accuracy Measure	β	SE	*t/z*	*p*
Intercept	0.162	0.197	0.82	
Length	0.337	0.087	3.86	***
Ambiguity	0.276	0.087	3.16	**
Age	0.322	0.259	1.24	
Length x Ambiguity	0.504	0.175	2.88	**
Length x Age	0.048	0.174	0.27	
Ambiguity x Age	0.045	0.174	0.26	
Length x Ambiguity x Age	0.041	0.348	0.12	

*0 '***' 0.001 '***' 0.01 '**' 0.05 '*'*

5.3.2. Eye movement results for the syntactic disambiguation region (e.g., *was slowly*)

5.3.2.1. Main effects

There was evidence of greater disruption to reading for ambiguous compared to unambiguous sentences in all measures, demonstrating a clear "garden path" effect (see Table 14 for *M* and *SE* and Table 15 for all GLM results). Furthermore, "long" sentences resulted in longer first-pass and regression path reading times, and more regressions in to the syntactic disambiguation region than "short" sentences. An age effect showed that older adults had longer first-pass and total reading times compared to their younger counterparts. In contrast, young adults had longer regression path reading times than older adults. Note that these main effects are reported for transparency but are qualified by interactions which are described in the next section.

5.3.2.2. Interactions

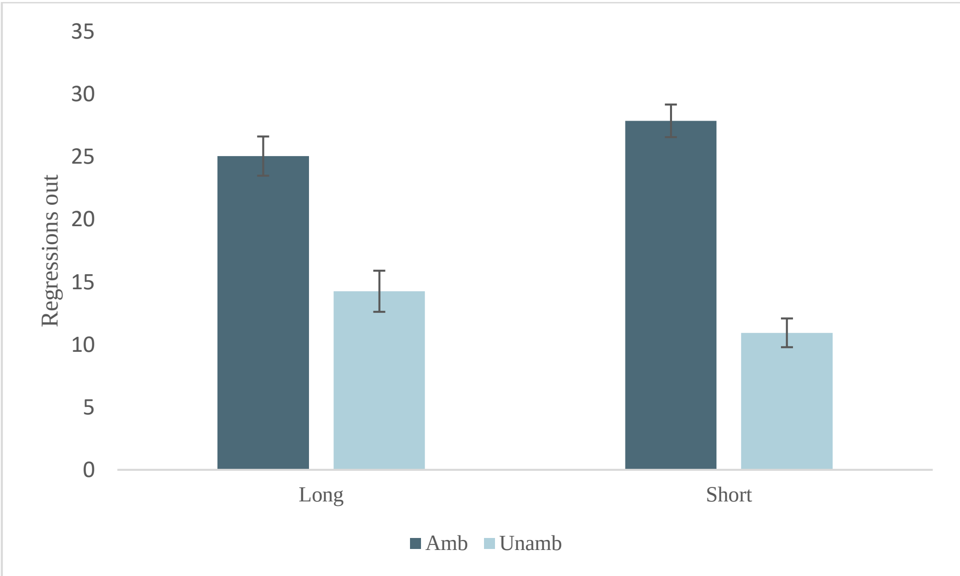
There were Length x Ambiguity interactions in first-pass, total reading times and regressions out of the syntactic disambiguation phrase, but comparisons simply reflected a main effect of ambiguity; longer reading times occurred for both long ambiguous (first-pass reading times = 465 msec; total reading times = 941 msec) than long unambiguous conditions (first-pass reading times = 388 msec; total reading times = 609 msec) and for short ambiguous (first-pass reading times = 452 msec; total reading times = 897 msec) and short unambiguous sentences (first-pass reading times = 365 msec; total reading times = 603 msec). For regressions out of the syntactic disambiguation phase, more regressions out were made for long ambiguous (0.25) than long unambiguous sentences (0.14) (see Figure 14).

Furthermore, more regressions out of the syntactic disambiguation region were made for short ambiguous sentences (0.28) than short unambiguous sentences (0.11). In the probability of fixating measure, the Length x Ambiguity interaction showed no differences between long ambiguous (0.92) and short ambiguous sentences (0.93), however long unambiguous sentences (0.92) were less likely to be skipped than short unambiguous sentences (0.89).

Overall, these results highlight readers' initial difficulty in processing ambiguous sentences, particularly when there was a short distance between the ambiguous information and disambiguating phrase.

Figure 14

Regressions out of the syntactic disambiguation region for length and ambiguity conditions.



Length x Ambiguity, Length x Age, and Ambiguity x Age interactions were found in regression path reading times and were qualified by a Length x Ambiguity x Age interaction, which showed that young adults had longer regression path reading times for "long" ambiguous sentences than older adults (see Figure 15). There were no other significant age differences.

Figure 15

Young and older adults' regression path reading times per condition at the syntactic disambiguation region

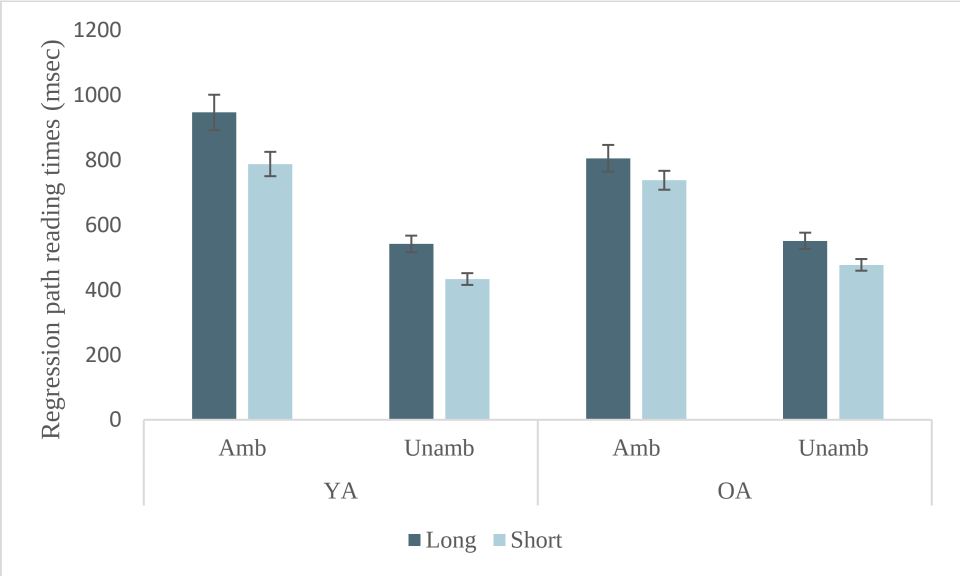


Table 14

Means and standard error for each reading measure by condition and age group.

Region	Age	Ambiguity	Length	First pass reading times	Regression path reading times	Total reading times	Skipping	RI	RO
Syntactic Disambiguation	Younger adults	Ambiguous	Long	455 (14)	947 (55)	958 (35)	92 (1)	30 (2)	28 (2)
			Short	434 (14)	788 (38)	850 (29)	93 (1)	33 (2)	28 (2)
		Unambiguous	Long	381 (12)	542 (26)	599 (21)	91 (1)	27 (2)	16 (2)
			Short	348 (9)	434 (18)	545 (18)	89 (2)	28 (2)	10 (2)
	Older adults	Ambiguous	Long	475 (13)	806 (41)	923 (30)	91 (1)	36 (2)	22 (2)
			Short	469 (12)	738 (29)	943 (29)	93 (1)	39 (3)	27 (2)
		Unambiguous	Long	395 (10)	551 (25)	619 (21)	92 (1)	25 (2)	12 (2)
			Short	383 (10)	477 (18)	660 (21)	89 (2)	26 (2)	12 (2)

Semantic Disambiguation	Younger adults	Ambiguous	Long	426 (13)	1104 (68)	739 (22)	99 (1)	11 (2)	34 (2)
			Short	411 (11)	897 (44)	728 (21)	100 (0)	19 (2)	33 (2)
		Unambiguous	Long	418 (12)	657 (34)	601 (17)	99 (0)	10 (2)	22 (2)
			Short	437 (11)	591 (22)	627 (17)	99 (0)	15 (2)	17 (2)
	Older adults	Ambiguous	Long	415 (11)	966 (51)	723 (17)	99 (0)	10 (2)	33 (2)
			Short	435 (12)	814 (36)	765 (21)	100 (0)	20 (2)	31 (2)
		Unambiguous	Long	435 (11)	608 (24)	649 (18)	100 (0)	14 (2)	19 (2)
			Short	431 (10)	665 (28)	693 (18)	100 (0)	21 (2)	24 (2)
Wrap-up	Younger adults	Ambiguous	Long	475 (16)	1039 (74)	638 (25)	96 (.92)	13 (2)	24 (2)
			Short	491 (18)	1003 (59)	646 (21)	97 (1)	7 (1)	26 (2)
		Unambiguous	Long	507 (17)	838 (41)	640 (23)	97 (1)	10 (2)	23 (2)
			Short	538 (18)	870 (40)	659 (22)	99 (1)	7 (1)	29 (2)
	Older adults	Ambiguous	Long	509 (15)	1238 (73)	703 (22)	98 (1)	14 (2)	39 (2)
			Short	573 (19)	1254 (64)	804 (27)	98 (1)	8 (1)	43 (3)
		Unambiguous	Long	556 (18)	969 (49)	713 (23)	99 (1)	11 (2)	32 (2)

Short	611 (20)	1084 (46)	774 (28)	96 (1)	8 (1)	41 (3)

Table 15

Linear and binomial mixed-effects models by segment and eye-movement measures.

	Syntactic disambiguation region				Semantic disambiguation region				Sentence wrap-up region			
Measure	β	SE	*t/z*	*p*	β	SE	*t/z*	*p*	β	SE	*t/z*	*p*
First-pass reading times												
Intercept	418.429	8.227	50.86	***	438.0378	9.206	47.58	***	556.235	10.791	51.54	***
Length	-19.888	8.947	2.22	*	10.4144	4.8561	2.15	*	33.639	6.204	5.42	***
Ambiguity	-77.417	7.323	10.57	***	9.3173	8.1844	1.14		42.127	8.334	5.06	***
Age	25.466	7.469	3.41	***	4.4156	10.1425	0.44		61.395	7.897	7.77	***
Length x Ambiguity	-24.79	7.544	3.29	**	-2.3403	8.3125	0.28		10.144	9.207	1.10	
Length x Age	13.113	8.649	1.52		-0.9067	9.642	0.09		30.046	10.583	2.84	**
Ambiguity x Age	-1.931	9.154	0.21		-4.1808	9.6274	0.43		-6.497	8.272	0.79	
Length x Ambiguity x Age	2.207	8.044	0.27		-33.7176	8.3627	4.03	***	-33.756	8.634	3.91	***

Regression path reading times												
Intercept	646.262	8.437	76.60	***	788.394	7.997	98.59	***	1026.472	8.251	124.40	***
Length	-62.738	8.040	7.80	***	-80.886	9.302	8.70	***	56.662	7.516	7.539	***
Ambiguity	-302.053	6.730	44.88	***	-297.754	8.903	33.44	***	170.953	6.915	24.720	***
Age	-27.096	9.122	2.97	**	-35.472	8.863	4.00	***	179.495	9.412	19.071	***
Length x Ambiguity	-31.689	9.804	3.23	**	136.36	8.964	15.21	***	18.603	9.915	1.876	
Length x Age	75.738	9.111	8.31	***	83.063	10.315	8.05	***	73.136	9.404	7.77	***
Ambiguity x Age	97.789	10.619	9.21	***	101.937	7.965	12.80	***	97.412	13.640	-1.80	
Length x Ambiguity x Age	-98.64	12.091	8.16	***	-0.347	12.592	0.03		9.753	10.271	-0.91	
Regressions in												
Intercept	-0.870	0.100	8.66	***	-2.070	0.166	-12.51	***	-2.924	0.229	-12.74	***
Length	0.246	0.084	2.93	**	0.647	0.111	5.83	***	-0.533	0.201	-2.65	**
Ambiguity	-0.267	0.089	3.01	**	-0.057	0.127	-0.44		-0.028	0.137	-0.21	

Age	0.193	0.171	1.13		0.122	0.173	0.71		0.087	0.228	0.38	
Length x Ambiguity	0.150	0.168	0.90		-0.231	0.222	-1.04		0.063	0.276	0.23	
Length x Age	0.275	0.168	1.64		0.152	0.222	0.69		0.086	0.272	0.32	
Ambiguity x Age	-0.183	0.173	1.06		0.459	0.237	1.94		-0.002	0.270	-0.01	
Length x Ambiguity x Age	0.453	0.335	1.35		-0.238	0.442	-0.54		0.149	0.543	0.27	
Regressions out												
Intercept	-1.784	0.154	-11.61	***	-1.201	0.113	-10.59	***	-0.926	0.141	-6.56	***
Length	0.082	0.185	0.44		-0.050	0.101	-0.50		0.302	0.087	3.46	***
Ambiguity	-0.900	0.139	-6.49	***	-0.727	0.121	-6.01	***	-0.111	0.094	-1.18	
Age	-0.197	0.171	-1.15		0.011	0.168	0.06		0.638	0.205	3.12	**
Length x Ambiguity	-0.464	0.210	-2.21	*	0.102	0.182	0.56		0.306	0.175	1.75	
Length x Age	0.483	0.208	2.32	*	0.304	0.180	1.69		0.125	0.174	0.72	
Ambiguity x Age	0.186	0.276	0.67		0.215	0.180	1.20		-0.313	0.188	-1.66	

Length x Ambiguity x Age	0.243	0.415	0.59		0.630	0.360	1.75		0.076	0.349	0.22	
Total reading times												
Intercept	772.006	9.602	80.40	***	709.865	13.157	53.95	***	748.394	9.48	78.94	***
Length	-11.694	9.031	1.30		33.99	17.398	1.95		49.965	11.434	4.37	***
Ambiguity	-322.067	8.304	38.79	***	-92.575	7.891	11.73	***	-3.155	7.834	-0.40	
Age	57.072	9.508	6.00	***	33.692	9.756	3.45	***	79.595	7.849	10.14	***
Length x Ambiguity	-17.703	7.889	2.24	*	19.96	13.935	1.43		-4.913	9.217	-0.53	
Length x Age	84.251	6.556	12.85	***	21.349	8.608	2.48	*	38.67	6.927	5.58	***
Ambiguity x Age	18.299	6.868	2.67	**	43.664	11.563	3.78	***	-12.11	6.53	-1.86	
Length x Ambiguity x Age	-19.424	10.095	1.92		3.877	9.645	0.40		-40.32	9.059	-4.45	***
Probability of skipping												
Intercept	3.995	0.372	10.73	***	9.980	22.024	0.45		5.127	0.387	13.26	
Length	0.194	0.249	0.78		4.764	43.830	0.11		0.023	0.273	0.08	
Ambiguity	-0.418	0.191	-2.19	*	4.189	43.827	0.10		0.062	0.419	0.15	

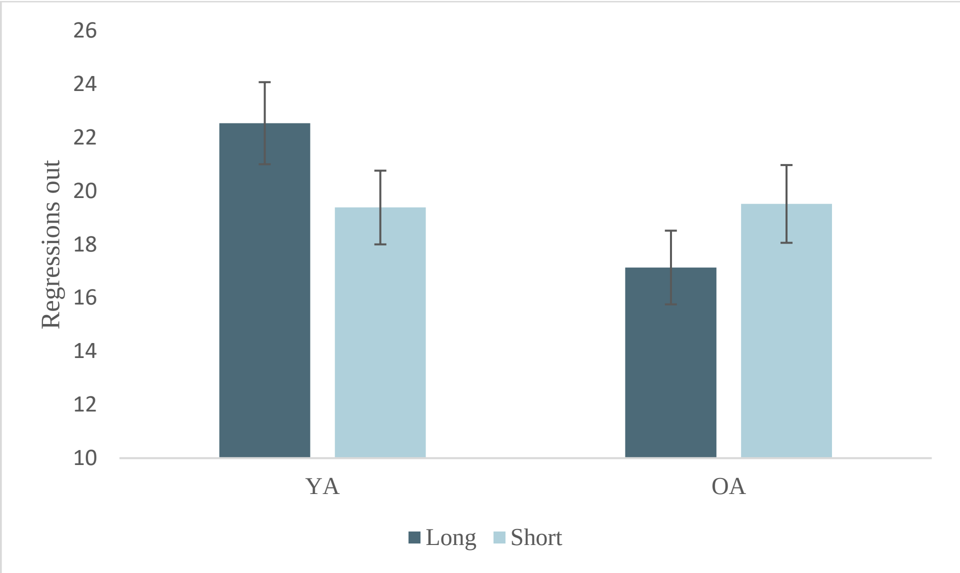
Age	-0.063	0.261	-0.24		4.777	43.832	0.11	0.304	0.464	0.65
Length x Ambiguity	-1.086	0.367	-2.96	**	7.339	87.654	0.08	-0.291	0.548	-0.53
Length x Age	-0.154	0.360	-0.43		8.089	87.654	0.09	-1.007	0.541	-1.86
Ambiguity x Age	0.147	0.371	0.40		9.246	87.655	0.11	-0.650	0.541	-1.20
Length x Ambiguity x Age	-0.394	0.717	-0.55		17.572	175.323	0.10	-1.816	1.084	-1.68

*0 '***' 0.001 '**' 0.01 '*' 0.05 '.'*

Similarly, a Length x Age interaction in regressions out showed no differences between young and older readers for "short" sentences, however, for "long" sentences, young adults made more regressions out than older adults (see Figure 16).

Figure 16

Young and older adults' regressions out of the syntactic disambiguation region for length condition



In total reading times, a Length x Age interaction indicated that older adults (804 msec) had longer reading times for "short" sentences than young adults (701 msec), but there was no age effect for "long" sentences (young adults = 786 msec; older adults = 774 msec). An Ambiguity x Age interaction indicated that older adults (639 msec) had longer total reading times for unambiguous sentences compared to young adults (572 msec) but there was no age difference for ambiguous sentences (young adults = 905 msec; older adults = 933 msec). These findings may generally reflect older adult readers' tendency to read more slowly than their younger counterparts. Interestingly, the ambiguity may have resulted in young adults also reading more slowly.

In sum, the results found in this region indicate a clear ambiguity effect. Furthermore, findings in later measures generally show that young adults dedicated more time to the reanalysis of “long” sentences, with this group showing longer regression path reading times and more regressions out of the region.

5.3.3. Eye movement results for the semantic disambiguation region (e.g., *being loaded*)

5.3.3.1. Main effects

There were longer regression path, longer total reading times, and more regressions out for ambiguous sentences compared to unambiguous controls, indicating a sustained “garden path” effect. “Long” sentences resulted in longer first-pass and regression path reading times, more regressions in, and marginally longer total reading times than “short” sentences. Older adults had longer total reading times than younger adults, but similarly to the syntactic disambiguation phrase, younger adults had longer regression path reading times than older adults. The next section describes interactions as the main effects are qualified by these.

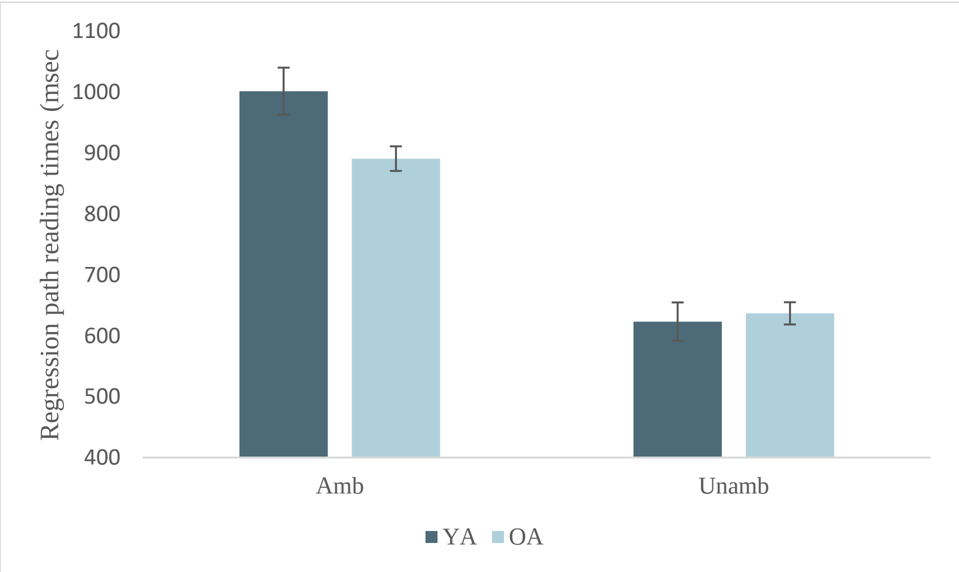
5.3.3.2. Interactions

There was a Length x Ambiguity x Age interaction in first-pass reading times revealing that older adults had longer first-pass reading times for “short” ambiguous sentences compared to “long” ambiguous sentences. There was no difference for older adults’ first-pass reading times for unambiguous sentences. On the other hand, there was no difference in young adults’ first-pass reading times for ambiguous sentences. However, in unambiguous sentences, young adults had longer first-pass reading times for “short” unambiguous sentences compared to “long” unambiguous sentences. The results may indicate that older adults are more likely to attempt to reanalyse short ambiguous sentences. However, when older adults hold on to the incorrect interpretation for longer, for instance, in the long ambiguous sentences, they are less inclined to reanalyse the initial interpretation.

A Length x Ambiguity interaction was found for regression path reading times. “Long” ambiguous sentences (1035 msec) resulted in significantly longer reading times than “short” ambiguous sentences (855 msec) but there were no length effects for unambiguous controls (long = 632 msec; short = 628 msec). This suggests that reanalysis of the initial ambiguity was still taking place during the semantic disambiguation region, especially for more complex sentences.

Figure 17

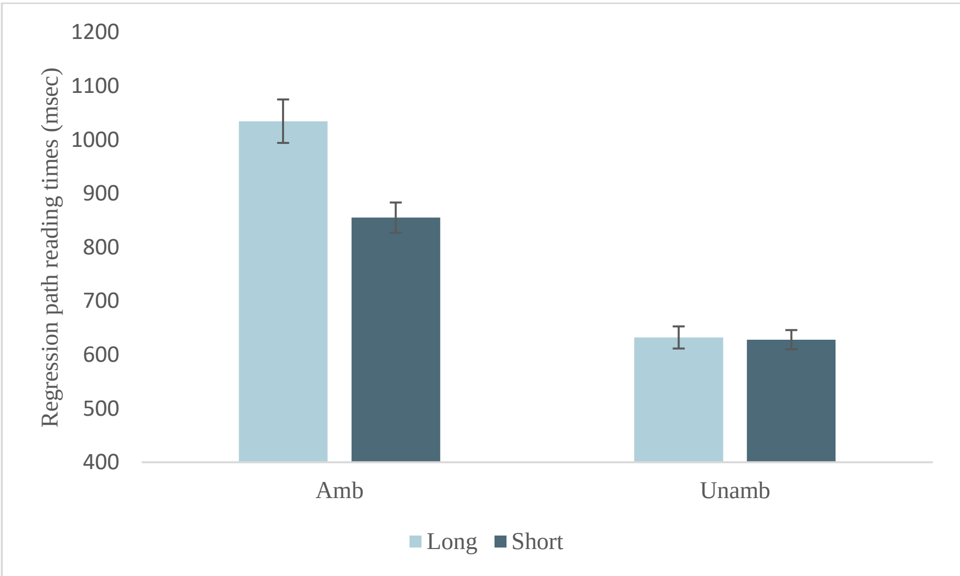
Young and older adults’ regression path reading times for ambiguity at the semantic disambiguation region



There was also a significant Age x Ambiguity interaction in regression path reading times (see Figure 17). Young adults had longer regression path reading times for ambiguous sentences compared to older adults, whereas there was no effect of age for unambiguous sentences. This interaction suggests that young adults spent more time on reanalysis than their older counterparts. A similar pattern was found with a Length x Age interaction (see Figure 18). Younger adults had longer regression path reading times for “short” sentences than older adults but there were no age differences for “long” sentences. These results suggest young adults spend more time re-inspecting previous parts of the text in order to overcome processing difficulty than older adults do.

Figure 18

Young and older adults’ regression path reading times for length at the semantic disambiguation region



There was a Length x Age interaction in total reading times (see Figure 19), showing no differences between long and short sentences for young adults. In contrast, older adults had longer total reading times for “short” sentences compared to “long” sentences. There was also an Ambiguity x Age interaction in total reading times (see Figure 20), showing that older adults’ total reading times were no different for ambiguous compared to unambiguous sentences. On the other hand, young adults had longer total reading times for ambiguous sentences compared to unambiguous sentences.

Figure 19

Young and older adults' total reading times for length at the semantic disambiguation region

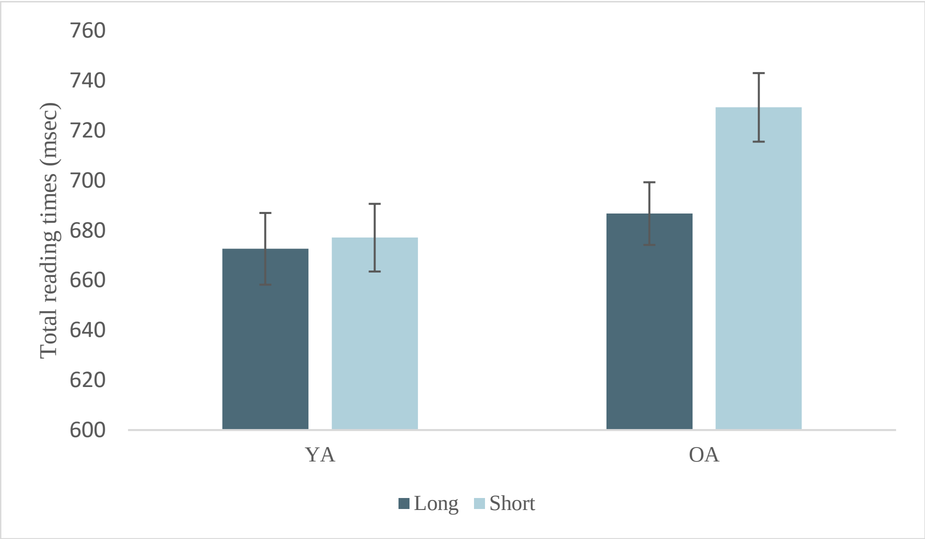
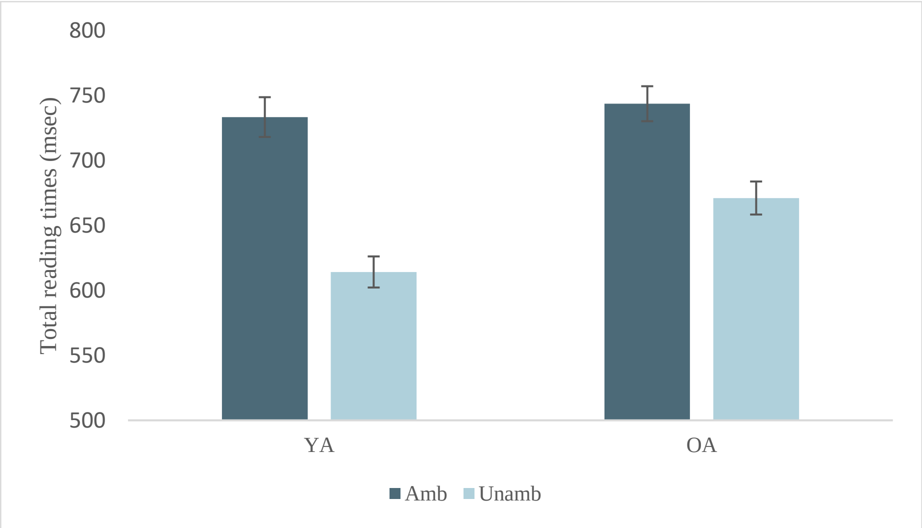


Figure 20

Young and older adults' total reading times for ambiguity at the semantic disambiguation region



5.3.4. Eye movement results for the wrap-up region (e.g., *in a garage nearby*)

5.3.4.1. Main effects

Ambiguous sentences resulted in shorter first-pass reading times but longer regression path reading times than unambiguous sentences. “Short” sentences resulted in greater disruption than “long” sentences in all eye movement measures except the probability of skipping. An age effect showed that older adults had longer first pass reading times, regression path reading times and made more regressions out of the wrap up region compared to young adults. These main effect results are in direct contrast to the results found in the previous regions where longer regression path reading times were found for long than short sentences and age effects showed young adults had longer regression path reading times than older adults. These results are somewhat expected as results illustrate the *wrap-up effect* (Rayner et al., 2000) where longer processing times are found at the end of sentence boundaries, especially for syntactic and semantic ambiguities (Hagoort, 2003; Payne et al., 2014) and especially in older adult readers (Payne & Stine-Morrow, 2014). Note that these effects are qualified by interactions which are described in the next section.

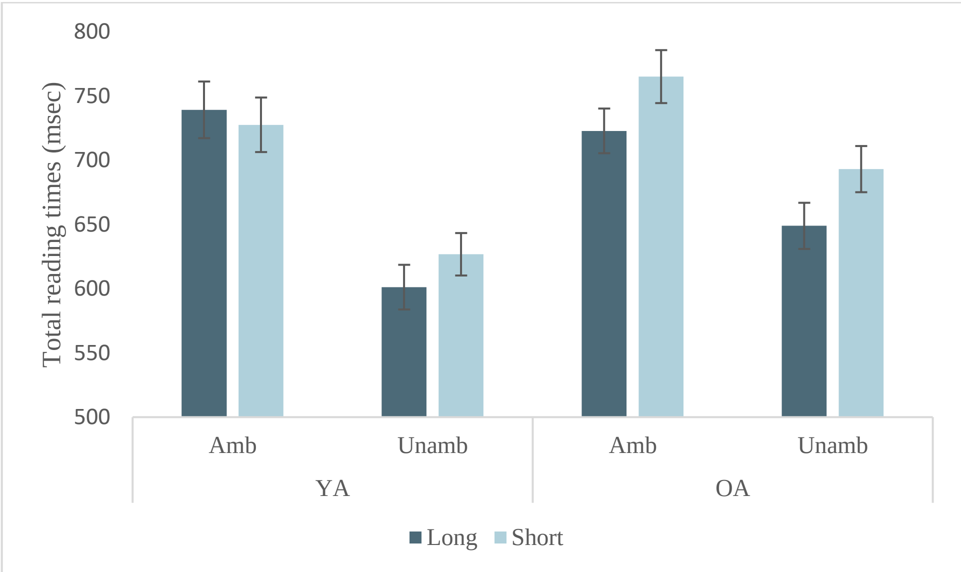
5.3.4.2. Interactions

There was a Length x Age interaction in regression path reading times showing that older adults had longer reading times for “short” (1170 msec) compared to “long” sentences (1105 msec), whereas there was no length effect in the young adult group (long = 941 msec; short = 936 msec). Further analysis of an Ambiguity x Age interaction simply reflected a main effect of age, showing that older adults had longer regression path reading times for both ambiguous (1246 msec) and unambiguous sentences (1026 msec) compared to their younger counterparts (ambiguous = 1021 msec; unambiguous = 854 msec).

Length x Age interactions in first-pass and total reading times were both qualified by Length x Ambiguity x Age interactions. First-pass and total reading times (see Figure 21) showed a similar pattern; older adults had longer reading times for ambiguous "short" (First-pass = 573 msec) sentences compared to ambiguous "long" sentences (First-pass = 509 msec). However, no differences in reading times were found for young adults processing ambiguous "long" (First-pass = 475 msec) versus "short" sentences (First-pass = 491 msec). Similar to results for the semantic disambiguation phrase, these results suggest that older adults attempt reanalysis of short sentences but are less likely to do so for long sentences.

Figure 21

Young and older adults' total reading times by condition at the wrap-up region



5.3.5. Is there a relationship between good enough processing and eye movement behaviour?

Another important issue to examine is the relationship between eye movement behaviour and ultimate comprehension. To the authors' knowledge, few studies have previously directly examined the relationship between eye movement behaviour and good enough processing. Wonnacott et al. (2016) examined children's eye movements while they read sentences containing temporary syntactic ambiguities. Interestingly, longer reading times on ambiguous sentences were not associated with a greater number of errors on comprehension questions that followed sentences containing temporary syntactic ambiguities. Instead, more regressions made into an earlier region of text was considered a positive predictor of accuracy (note that this was the case for both unambiguous and ambiguous sentences). A further eye-tracking study conducted by Huang and Ferreira (2021) found longer first-pass and total reading times predicted more errors in comprehending garden-path sentences. We aimed to examine the relationship between eye movements and comprehension in young and older adults.

To explore the relationship between eye movement behaviour and good enough processing, a logit mixed model was conducted. Accuracy scores and all eye-movement measures were inputted into a logit model (see Table 16). More regressions out of the syntatic disambiguation region (estimate = .003, *SE* = .001, z = 3.37), semantic disambiguation region (estimate = .002, *SE* = .001, z = 2.33), and sentence wrap-up region (estimate = .002, *SE* = .001, z = 2.26) were significant predictors of correctly answering comprehension questions. Increased accuracy was also associated with a lower probability of skipping the semantic disambiguation phrase (estimate = -.018, *SE* = .008, z = -2.33) and wrap-up regions (estimate = -.005, *SE* = .002, z = -2.01). Furthermore, shorter first pass reading times (estimate = .000, *SE* = .000, z = -3.36) and more regressions in to the sentence wrap-up region (estimate = .004, *SE* = .001, z = 3.43) predicted answering comprehension questions correctly.

Table 16

Logit-model scores by age group, condition and segment.

Accuracy Measure	Reading measure	β	SE	*t/z*	*p*
Syntactic disambiguation region	Intercept	0.177	0.068	2.60	**
	First pass reading times	0.000	0.000	-1.24	
	Intercept	0.070	0.051	1.37	
	Regression path reading times	0.000	0.000	1.07	
	Intercept	0.105	0.026	1.67	
	Total reading times	0.000	0.000	0.03	
	Intercept	0.074	0.044	1.66	
	Regressions in	0.001	0.001	1.30	
	Intercept	0.045	0.041	1.10	
	Regressions out	0.003	0.001	3.37	***
	Intercept	0.031	0.126	0.25	
	Skipping	0.001	0.001	0.62	
Semantic disambiguation region	Intercept	0.181	0.080	2.28	*
	First pass reading times	0.000	0.000	-1.07	
	Intercept	0.068	0.052	1.32	
	Regression path reading times	0.000	0.000	1.05	
	Intercept	0.133	0.077	1.72	
	Total reading times	0.000	0.000	0.39	
	Intercept	0.079	0.040	1.97	*
	Regressions in	0.002	0.001	1.79	

	Intercept	0.055	0.043	1.28	
	Regressions out	0.002	0.001	2.33	*
	Intercept	1.872	0.759	2.47	*
	Skipping	-0.018	0.008	-2.33	*
Wrap up region	Intercept	0.290	0.066	4.40	***
	First pass reading times	0.000	0.000	-3.36	***
	Intercept	0.151	0.050	3.04	**
	Regression path reading times	-0.000	0.000	-1.34	
	Intercept	0.204	0.066	3.11	**
	Total reading times	-0.000	0.000	-1.80	
	Intercept	0.064	0.039	1.66	
	Regressions in	0.004	0.001	3.43	***
	Intercept	0.049	0.044	1.11	
	Regressions out	0.002	0.001	2.26	*
	Intercept	0.592	0.244	2.42	*
	Skipping	-0.005	0.002	-2.01	*

*0.01 ‘***’ 0.01 ‘**’ 0.05 ‘*’*

5.4. Discussion

The present study tracked eye movements during reading to investigate a number of factors in the online processing and comprehension of temporarily ambiguous sentences, including age, length of the sentence, and working memory. The relationship between readers' eye movement behaviour and comprehension performance was also examined. No significant age differences were found in the comprehension data, however, there were a number of age differences in the eye movement data. The eye movement data clearly showed that both young and older adult readers were 'garden-pathed' by sentences containing temporary syntactic ambiguities. Another aim was to examine whether working memory mediates age differences in good enough processing. To do this the distance between the ambiguity and the disambiguating information was manipulated and a direct working memory assessment was completed by participants. The eye movement data showed increased processing difficulty was found when the distance between the ambiguous phrase and subsequent disambiguating information was longer. Good enough processing was more prevalent for long ambiguous sentences compared to short ambiguous sentences, as evidenced in a lower number of correct responses to comprehension questions. The direct assessment of working memory did not converge in the GLM analyses or was not deemed a better model than the base model. A final aim of the study was to explore the relationship between readers' eye movements and comprehension accuracy. This analysis showed that first pass reading times, the probability of skipping, the number of regressions made in and out of regions were predictors of correctly responding to comprehension questions. These findings are discussed further in the following sections.

5.4.1. Comprehension accuracy

We predicted that older adults would be more likely to engage in good enough processing, resulting in a lower proportion of correct answers to comprehension questions than their younger counterparts (Christianson et al., 2006, Experiment 1a and b; Dede, 2014; Kemper et al., 2004). However, despite a numerical trend towards older adults having a lower proportion of correct answers, no significant age differences were found. These results are commensurate with some self-paced reading studies (Christianson et al., 2006, Experiment 2) and sentence picture matching tasks (Malyutina & den Ouden, 2016) that have used similar materials and found no significant differences in comprehension between young and older adult readers. Our results are also in line with more general reading literature, which suggests that while older adults experience greater processing difficulty with complex sentences, their comprehension remains unaffected compared to their younger counterparts (Kliegl et al., 2004; Obler et al., 1991). Discrepancies between studies that identified age differences in processing of sentences containing temporary syntactic ambiguities and the present findings may lie in the different types of ambiguous sentences used.

Previous self-paced reading studies have found that when the distance between the ambiguous phrase and the disambiguating information is longer, young adults are more likely to make errors in comprehension questions than when the distance is shorter (Christianson et al., 2001; Ferreira & Henderson, 1991). In accordance with these previous experiments, the present findings indicated that “long” ambiguous sentences resulted in a higher proportion of incorrect answers compared to “short” ambiguous sentences; there was no length effect for unambiguous controls. There are different explanations as to why comprehension accuracy is affected by length. One possibility is that the longer the initial (mis)representation is held for, the less likely the reader is to reanalyse this misinterpretation. An alternative explanation is that the additional words create interference in the retrieval of information.

The current results support previous findings of a lingering misinterpretation (Christianson et al., 2001; Jacob & Felser, 2016; Sturt, 2007). A lingering misinterpretation occurs when initial semantic commitments (e.g., *the gentleman eating the burger*) persist even after reanalysis (Christianson et al., 2001; Jacob & Felser, 2016; Slattery et al., 2013; Sturt, 2007). The sentences used in the present experiment include both syntactically contrasting information (*were still*) and semantically contrasting information (being *reheated*). A number of explanations have been put forward as to why lingering misinterpretation occurs. One reason may be a failure in fully erasing the initial misinterpretation (Slattery et al., 2013a). Alternatively, there may be a failure of the processing system to combine incoming syntactic and semantic information successfully (Sturt, 2007), or an error in the memory system (Kaschak & Glenberg, 2004). Further research is needed to distinguish these explanations.

5.4.2. The relationship between eye movement behaviour and good enough processing

To obtain a comprehensive understanding of the processes underlying language comprehension the relationship between online processing and offline comprehension was explored (Ferreira et al., 2019). Eye movements were used as an index of online processing and accuracy of comprehension questions were an index of offline comprehension. In Experiment 1, anomaly detection accuracy and eye movement measures were inputted into a logistic regression. At the pre-critical region, longer total reading times and more regressions predicted greater accuracy, as did more regressions into the critical region. At the post-critical region, shorter first-pass reading times predicted more accurate detection.

Similar to Experiment 1, accuracy on comprehension questions and eye movement measures were inputted into a logistic regression. The present findings showed that more regressions out of all regions, more regressions in to the wrap-up region, shorter first pass reading times in the wrap-up region and a lower probability of skipping the semantic disambiguation phrase and wrap-up region were associated with better performance on comprehension questions.

Previously, Huang and Ferreira (2021) examined eye movement behaviour and offline comprehension and longer first-pass and total reading times predicted more errors in comprehending garden-path sentences. No effect was found for go-past time and regressions were not measured in this experiment. Wonnacott et al. (2016) measured children's eye movements while they read sentences containing temporary syntactic ambiguities, as well as assessing their performance on comprehension questions. Interestingly, longer reading times did not predict accuracy, which is consistent with what was found in the present study examining adult readers. Instead, and also consistent with the present study, Wonnacott et al. found that more regressions made into an earlier region of text was considered a significant predictor of accuracy in both ambiguous and unambiguous sentences. This finding is compatible with previous research that has shown that regressions indicate mindful reading (Schroeder, 2011). Thus, readers who engage in good enough processing would be less likely to make regressions and as a consequence more likely to have a lingering misinterpretation.

Furthermore, results from the present study showed that less skipping of the semantic disambiguating and wrap-up regions were related to higher accuracy rates. More recently, in addition to visual factors, linguistic factors, such as plausibility, have been found to influence word skipping. Skipping has also been linked to more experienced readers (Faber et al., 2020), older adults (McGowan, 2014; McGowan et al., 2015; Paterson et al., 2013; Rayner et al., 2006, 2011) and better spellers (Drieghe et al., 2019). Results from the present study indicate that less word skipping may be related to more accurate processing of syntactically complex sentences.

5.4.3. Eye movement data

As predicted, the present study showed that both young and older adults were initially "garden-pathed" by temporarily ambiguous sentences, with longer reading times and more regressions observed for ambiguous than unambiguous materials. This supports previous studies that have found the same pattern for young and older adults (Kemper et al., 2004) and native and non-native readers (Jacob & Felser, 2016). Similarly to the native readers in Jacob and Felser's (2016) study, young adults were more likely to attempt to integrate the auxiliary (*were still*) with the noun phrase (*While the gentleman was eating the burgers*) as this group had longer regression path reading times for ambiguous sentences than older adult readers.

In the present study, despite young and older adults making more regressions and having longer regression path reading times and total reading times for ambiguous sentences at the syntactic and semantic disambiguation phrase, a lingering misinterpretation still persisted as comprehension rates were low. Furthermore, older adults were less likely to reanalyse the ambiguity compared to younger adults at the syntactic and semantic disambiguating regions; instead, older adults had longer regression path reading times than younger adults at the sentence wrap-up region.

In relation to effects of sentence length, previous research using similar stimuli has shown longer regression path reading times for ambiguous "short" sentences compared to ambiguous "long" sentences (Jacob & Felser, 2016). In the current experiment, older adults were found to experience greater difficulty in processing "short" than "long" sentences. In general, young adults were more sensitive to length effects, making more effort to reconcile disruption at the syntactic disambiguation region. Specifically, young adults regressed back to earlier parts of the text for "long" sentences in both disambiguating regions, which meant that relatively little extra processing was required at the wrap-up region. Importantly, these findings illustrate the different strategies young and older adults used to disambiguate sentences. Specifically, older adults spent more time resolving processing difficulties when encountering the sentence wrap-up region, however, younger adults spent longer regressing back to earlier parts of the sentences to attempt to immediately resolve the ambiguity. The results suggest that a memory component may be required for successful reanalysis when processing temporary syntactic ambiguities as older adults were less likely to attempt to reanalyse long sentences. In contrast, young adults were able to reanalyse sentences immediately, however further investigation of this is necessary as a direct assessment of working memory did not converge in LMEM models.

5.4.4. Conclusion

In Chapter 5, eye movements during reading were used to investigate a number of factors affecting the processing and interpretation of temporarily ambiguous sentences with a focus on age, length of sentence, and working memory. Eye movement data clearly showed that young and older adults used different strategies to resolve temporary syntactic ambiguities. Younger adults used an immediate approach, immediately attempting reanalysis on encountering information which is inconsistent with their previous (mis)interpretation of temporarily ambiguous information. In contrast, older adults were more likely to wait until the sentence wrap-up region to reanalyse information within the sentence. This study also furthered our understanding of the relationship between eye movement behaviour and comprehension accuracy, showing that more regressive eye movements are associated with better comprehension accuracy. The findings of this study are consistent with the good enough processing framework showing that readers often process sentences at a shallow and incomplete level. Practically, results from this study demonstrate that comprehension remains intact into older adulthood but older adults may require additional time to process complex text.

In Chapter 6, Experiment 3 aimed to investigate the age differences in the relationship between online processing and offline comprehension. Older adults have been found to make longer forward eye movements (McGowan et al., 2015; McGowan et al., 2014; Paterson et al., 2013a; Rayner et al., 2006, 2011, 2013) and skip words more frequently (McGowan et al., 2014, 2015; Paterson et al., 2013; Rayner, et al., 2006, 2011, 2013) but then look back to words more often when their inferences are incorrect (Kliegl et al., 2004; McGowan et al., 2014, 2015; Paterson et al., 2013; Rayner et al., 2006, 2011, 2013). Rayner et al. (2006) suggested that older adults employed a risky reading strategy whereby they may try to guess the upcoming identities of words but then when these interpretations are incorrect, they are then more likely to regress back to earlier parts of the text. Furthermore, the good enough processing framework suggests adults utilise their world knowledge and experience employing heuristics during processing.

The consequences of risky reading behaviour for successful text comprehension are currently unknown and it was predicted that risky readers may be more likely to engage in good enough processing. To examine this, Experiment 3 was conducted to explore the relationship between the characteristics of risky reading and offline comprehension. The study aimed to replicate findings from a classic experiment that found evidence of good enough processing using non-canonical sentences (Ferreira, 2003). Specifically, plausible and implausible sentences were presented in an active and passive sentence structure and more errors were made in implausible passive sentences. Further analyses were then conducted to determine the relationship between characteristics of risky reading and good enough processing.

Chapter 6: Investigating the relationship between eye movements and ultimate comprehension

6.1. Introduction

Traditional models of language comprehension assume that the main goal of reading is to derive a complete and accurate representation of the text by combining every word of a sentence into a coherent syntactic structure (e.g., Chomsky, 1965; Ferreira & Clifton, 1986; Frazier & Rayner, 1982; MacDonald et al., 1994; Tanenhaus et al., 1995). In contrast, alternative theoretical approaches, such as shallow processing (Sanford & Sturt, 2002) and good enough processing frameworks (Ferreira et al., 2002) suggest that readers may instead develop a representation that relies on fast and frugal heuristics that may not be complete but is good enough for the situation presented. While there is a relatively large literature investigating this phenomenon, currently little is known about individual differences in depth of processing, or the relationship between 'on-line' processing and comprehension accuracy. Thus, the current study had three aims; to investigate whether there are age differences in good enough processing; to examine whether young and older adults display different patterns of reading behaviour when processing sentences which may result in 'good enough' interpretations; and to establish whether there is a relationship between eye movements during reading and 'good enough' processing.

As studied in the previous chapter, one example of a good enough interpretation that can occur is when readers process temporarily ambiguous sentences, such as *While the gentleman was eating the burgers were still being reheated in the microwave*. Initially, the phrase *the burgers* is often processed as the direct object of *eating* but when the reader encounters structurally inconsistent (*were*) and semantically inconsistent information (*being reheated*) *the burgers* should be reprocessed as the subject of the main clause (*in the* *microwave*). If readers subsequently answer ‘No’ to comprehension questions such as *Was the gentleman eating the burgers?* this suggests that they have ultimately correctly interpreted the sentence. If, however, readers answer ‘Yes’, this implies the initial misanalysis of the ambiguity has not been rejected, and has been taken as evidence for a phenomenon known as “lingering misinterpretation” (Christianson et al., 2001; Jacob & Felser, 2016; Slattery et al., 2013a; Sturt, 2007).

Interestingly, readers have also been shown to adopt shallow or good enough representations of text that is not ambiguous. For example, as studied in Chapter 4, when sentences contain an anomalous word or phrase, such as, *When an airplane crashes, where should the survivors be buried?* readers often fail to notice the anomalous word (*survivors*), despite knowledge of the meaning of *survivors*. This ‘semantic illusion’ illustrates that readers do not always process the core meaning of all the words within a sentence, instead individuals sometimes rely on context and their world knowledge. Thus, when the anomalous word (e.g., *survivors*) has a good global fit to the context (e.g., *an airplane crash*) the anomaly often goes undetected (Barton & Sanford, 1993).

Implausible sentences presented in a passive structure, such as, *The dog was bitten by the man*, are another case in which readers often adopt a ‘good enough’ interpretation. When asked to report “Who was the do-er?”, participants sometimes erroneously report that the dog bit the man (as our experience of the world would normally suggest), despite the sentence in fact stating that it was the man who bit the dog (Ferreira, 2003). In this example, readers were shown to derive a representation based on fast and frugal heuristics, that is, relying on world knowledge and a Subject-Verb-Object (SVO) bias rather than conducting a full parse. The above examples indicate that even for unambiguous sentences, the reader does not always create a veridical interpretation of text using algorithmic processing, instead, readers may utilise heuristics to derive a good enough representation.

Researchers have traditionally focused on the proportion of correct answers to comprehension questions as a measure of good enough processing (Christianson et al., 2001, 2006; Ferreira, 2003; Malyutina & den Ouden, 2016). However, monitoring the relationship between online processing and these offline measures of interpretation is particularly important to better understand the mechanisms underlying good enough processing (Ferreira et al., 2019). A number of studies have investigated this relationship using self-paced reading tasks. For example, across three self-paced reading experiments examining the processing of sentences containing temporary syntactic ambiguities, Christianson and Luke (2011) found that longer reading times were not associated with more accurate answers to comprehension questions. The authors noted that these results were surprising, especially since longer reading times on a word or phrase are seen as reliable signals of processing difficulty and implicit indications of comprehension (Staub & Rayner, 2007; Sturt, 2007). Similarly to Christianson and Luke (2011), Qian et al. (2018) conducted two self-paced reading experiments and also found no relationship between reading times and comprehension accuracy. Furthermore, Qian et al. conducted an event-related potential (ERP) study where sentences containing temporary syntactic ambiguities were presented word-by-word to participants. They found that P600 amplitude (which is often taken as an online index of syntactic processing) was not associated with performance on offline measures of comprehension.

Although these studies found no observable association between measures of online processing and offline measures of comprehension, it is important to note that participants are not able to read as they would naturally in self-paced reading tasks or tasks which involve word-by-word presentation of the text. In contrast, eye-tracking experiments are a method used to monitor readers' eye movements while they are reading text on a computer screen, allowing for common behaviours such as skipping over words and looking back at previous parts of the text. To the authors' knowledge, few studies have used eye-tracking methods to directly examine the relationship between online processing and offline comprehension. One example is a study conducted by Wonnacott et al. (2016), who presented sentences containing temporary syntactic ambiguities to young adults and children while their eye movements were monitored. In contrast to their predictions, longer reading times on ambiguous sentences were not associated with a greater number of errors in comprehension. Furthermore, more regressions positively predicted accuracy, however the regressions were not particularly targeted as this was the case for both unambiguous and ambiguous sentences.

More recently, Huang and Ferreira (2021) recorded participants eye movements whilst reading sentences containing temporarily ambiguous sentences and found that longer first-pass and total reading times predicted more errors in comprehending garden-path sentences. No effect was found for go-past time. Thus, it is clear that further investigation is warranted into the relationship between eye movements during reading and offline measures of interpretation (Ferreira et al., 2019).

A particularly interesting group to investigate good enough processing in is older adults. Older readers have been found to experience greater difficulty in processing text (Rayner, Reichle, et al., 2006), especially when it is syntactically complex (Obler et al., 1991). Previous research has shown that world knowledge typically increases across the lifespan (Schaie, 1994) and older adults often perform better on vocabulary tests than their younger counterparts (Ben-David et al., 2015; Keuleers et al., 2015). Coupled together, older adults may compensate for any difficulty in processing text by utilising their world and linguistic knowledge to make more experience-based inferences rather than engaging in algorithmic processing. That is, older adults may be more likely to engage in good enough processing.

Indeed, empirical studies investigating the eye movements of young and older readers have found evidence to suggest that older adults may employ a strategy where they try to use the context and their knowledge of the world to infer the identities of upcoming words. Specifically, when reading single sentences, older adults have been found to skip words more often (McGowan, 2014; McGowan et al., 2015; Paterson et al., 2013a; Rayner et al., 2006; 2011), regress back to earlier parts of text (McGowan et al., 2014, 2015; Paterson et al., 2013b; Rayner et al., 2006; 2011) and make longer forward eye movements (McGowan, 2014; McGowan et al., 2015; Paterson et al., 2013; Rayner et al., 2006; 2011) compared to their younger counterparts. Rayner et al. (2006) suggested that older adults employ a risky reading strategy whereby they may try to guess the upcoming identities of words but then when these interpretations are incorrect, they are then more likely to regress back to earlier parts of the text. Importantly, the consequences of risky reading behaviour for successful text comprehension are currently unknown. Therefore, it will be informative to assess whether there is a relationship between risky reading and good enough processing.

6.1.1. Present study and Predictions

The current experiment has three aims; to investigate whether young and older adults engage in good enough processing to the same extent; to examine whether older adults adopt risky reading in more complex texts than have been previously studied; to examine whether there is a relationship between risky reading and good enough processing.

To address these aims, an eye-tracking study was conducted in which the eye movements of young and older adults were monitored while they read active or passive sentences describing either plausible or implausible events such as (1a-1d):

1a) The knight protected the hopeful and loyal residents for as long as possible but the enemy was too strong. (*active plausible*).

1b) The residents protected the hopeful and loyal knight for as long as possible but the enemy was too strong (*active implausible*).

1c) The residents were protected by the hopeful and loyal knight for as long as possible but the enemy was too strong. (*passive plausible*).

1d) The knight was protected by the hopeful and loyal residents for as long as possible but the enemy was too strong (*passive implausible*).

After each item, a comprehension question was asked to assess depth of processing, such as, *Who did the protecting?* The relationship between participants' performance on comprehension questions and their eye movements while reading the sentences was assessed.

In terms of predictions, if readers engage in good enough processing, more errors on comprehension questions are expected for implausible sentences compared to plausible sentences, and for passive sentences compared to active sentences, as well as an interaction showing the most comprehension errors for implausible passive sentences (Ferreira, 2003). If older adults engage in a greater degree of good enough processing as a result of relying more on context and world knowledge, they should have lower accuracy scores than young adults.

If older adults adopt a risky reading strategy, older readers will make more regressions back to earlier parts of the text (McGowan et al., 2014, 2015; Paterson et al., 2013b; Rayner et al., 2006; 2011), be more likely to skip words (McGowan, 2014; McGowan et al., 2015; Paterson et al., 2013; Rayner et al., 2006; 2011) and make longer forward eye movements (McGowan, 2014; McGowan et al., 2015; Paterson et al., 2013; Rayner et al., 2006; 2011) than young adults.

Previous research has suggested that readers who are more likely to make regressions back to earlier parts of the text engage in less good enough processing (Wonnacott et al., 2016). In contrast to this, we might predict that if risky reading results in more shallow processing of the text, then increased risky reading behaviour (i.e., more word skipping, longer forward saccades, and more regression) should be associated with lower performance on comprehension questions, that is, with a greater degree of 'good enough' processing.

6.2. Method

6.2.1. Participants

The participant characteristics are the same as Experiment 1 as these experiments were conducted in the same session.

6.2.2. Design and stimuli

A 2 (Structure: Active x Passive) x 2 (*Plausibility*: Plausible x Implausible) x 2 (Age: Young x Older adults) mixed design was employed. Structure and Plausibility were within-subjects and Age group was between-subjects. There were 36 experimental items (see Appendix – Item A). All target sentences were preceded by a context sentence, for example, *The battle had lasted for hours.* Experimental sentences (see Table 17), were presented in four conditions, the sentence was either plausible (17a and 17c) or implausible (17b and 17d) and was presented in an active (17a and 17b) or passive sentence structure (17c and 17d). Items were presented across multiple lines. After each item, a comprehension question referring to the 'doer' was presented, for example, *Who did the protecting?* Followed by two possible answers *knight/residents*. Correct answers to the comprehension questions were used as an index of good enough processing. There were four counterbalanced presentation lists, resulting in participants seeing eight items in each of the four conditions. Stimuli were presented in a randomised order using a Latin Square. Thirty-six fillers were constructed, half were passive sentences and half were active sentences and all were plausible (see Appendix – Item B).

Table 17

Example experimental item in all conditions.

	Condition		Example Material
	Structure	Plausibility	
a.	Active	Plausible	The battle had lasted for hours. The knight protected the hopeful and loyal residents for as long as possible but the enemy was too strong.
b.	Active	Implausible	The battle had lasted for hours. The residents protected the hopeful and loyal knight for as long as possible but the enemy was too strong.
c.	Passive	Plausible	The battle had lasted for hours. The residents were protected by the hopeful and loyal knight for as long as possible but the enemy was too strong.
d.	Passive	Implausible	The battle had lasted for hours. The knight was protected by the hopeful and loyal residents for as long as possible but the enemy was too strong.

Plausibility pre-test:

A norming task was undertaken prior to the main experiment to assess the success of the plausibility manipulation. Participants consisted of 34 older adults (65 years old and above) and 38 younger adults (18-30 years old), the sample included 52 females and 18 males. Participants read 48 scenarios that were constructed by the author (e.g., *The knight protected the hopeful and loyal residents*) and were asked to rate plausibility using a 4-point Likert scale (very plausible to very implausible). Four lists were created with the structure and plausibility of each scenario balanced so each participant saw one version of the item. Item order was randomised in Qualtrics. Participants completed this task online in their own time which took approximately 20 minutes. Upon completion, participants could opt-in to a prize draw to win a £25 Amazon voucher. Sentences were omitted from the main eye-tracking experiment if the average rating from all participants for a plausible sentence was 'very implausible' or 'implausible' and for an implausible sentence if the average rating was 'very plausible' or 'plausible'. This resulted in 12 sentences being omitted (Appendix – Item C), and the remaining 36 items were the experimental items used for the main eye-tracking experiment.

6.2.3. Apparatus

An SR Research Ltd. EyeLink 1000 tower mounted eye-tracker was used to record gaze location every millisecond. The display screen resolution was 1024 x 768 and the refresh rate was 120Hz. Only the right eye was tracked although viewing was binocular. Participants viewed items on a 20-inch monitor at a viewing distance of 80 cm with three letters displayed per degree of visual angle. Sentences were displayed in Courier New 14-

point font as black text with sentences double spaced and presented on a white background. A chin and forehead rest were used to minimise movements.

6.2.4. Procedure

The study received ethical approval (Ref number: *14100-TK235-Neuroscience, Psychology and Behaviour*) from The University of Leicester Ethics Committee (Psychology). At the beginning of the experiment, participants completed the visual acuity tasks. Before completing the eye-tracking experiment, participants were informed that they should read the sentences for comprehension after which a question would be displayed. Prior to the presentation of the trials, a nine-point calibration procedure was conducted. Each trial then began with the presentation of a fixation cross. Once the participant was correctly fixating the cross, the item was automatically presented with the first letter of the sentence replacing the cross. Participants were instructed to press a button on a gamepad once they had finished reading the item and to press the left or right button when answering the comprehension question, which was always asking who performed the action described in the sentence (i.e., *Who did the protecting? knight/residents*). After the eye-tracking task, working memory and vocabulary tests were administered.

6.2.5. Analyses

The “clean” function in DataViewer (SR Research) was used to trim the data. Fixations shorter than 80ms, and which were located within one character space of the next or previous fixation, were merged into that nearby fixation, fixations shorter than 80ms or longer than 1200ms were discarded. Six items were removed from the final analyses due to experimenter error. The following eye movement measures were analysed as indices of risky reading: The *probability of fixating* is the likelihood that the target word is not by-passed on first-pass reading (skipped = 1); *Number of regressions* is the average number of backward eye movements made across the target sentence*; Progressive saccade length* is the average character length of forward eye movements made in the target sentence. Errors on the comprehension question data were scored as '1' and these scores were used as an index of good enough processing.

6.3. Results

A logistic regression was employed using the lme4 package (Baayen et al., 2008; Bates et al., 2015) to analyse data in R (Version 3.5.1; R Core Team, 2019). The binomial family and 'logit' link were utilised. Models included Age group (young vs. older adults), Structure (active vs. passive) and Plausibility (plausible vs. implausible) as fixed factors. Each eye movement measure when entered as a predictor variable was also a fixed factor when assessing accuracy. When assessing risky reading, each eye movement measure was the outcome variable and Age a fixed factor. Reported results are based on models that successfully converged (see Appendix – Item D for full syntax of reported models). Absolute *z*-values were considered significant if ≥ 1.96 for all models.

6.3.1. Risky reading

Initially, a logistic regression was conducted to examine whether older adult readers adopted a risky reading strategy compared to young readers. The model included Age group and was conducted separately for each eye movement measure; probability of skipping, progressive saccade length and regressive saccades (see Table 18 for *M* and *SE*). If older adults engaged in risky reading, then they were expected to make more regressive saccades, be more likely to skip words, and have longer progressive saccade lengths than young adults. This was indeed what the data showed (see Table 19), suggesting that older adults had adopted a 'riskier' reading strategy than young adults.

Table 18

Means and Standard error of errors per eye movement measure by age

Eye movement	Age	
	Young	Older
Progressive saccade length	2.53 (0.00)	2.91 (0.00)
Regressive saccade	11 (0.07)	12 (0.06)
Probability of skipping	8 (0.03)	9 (0.03)

Table 19

GLM model of eye movement measures and Age

	Estimate	Standard error	Z Value	
Progressive saccade length				
Intercept	2.72	0.00	921.35	***
Age	0.38	0.01	64.78	***
Regressive saccade				
Intercept	11.932	0.049	246.03	***
Age	0.262	0.097	2.71	***
Probability of skipping				
Intercept	8.95	0.02	447.08	***
Age	0.48	0.04	12.04	***

*0 '***' 0.001 '***' 0.01 '**' 0.05 '*'*

6.3.2. Shallow processing

The next analysis was conducted to examine whether participants engaged in good enough processing. A logistic regression was conducted between Structure x Plausibility x Age group to assess the influence of each of these factors on accuracy of responses to comprehension questions. Main effects of Structure and Plausibility were found (see Table 20 for *M* and *SE*). Passive structures resulted in more errors than active structures, and implausible sentences resulted in more errors than their plausible counterparts. Counter to our predictions, there was no significant interaction between Structure x Plausibility (see Table 21).

We predicted that older adults may be more likely to engage in good enough processing compared to young adult readers. In contrast to these predictions, older adults made fewer errors (M = 0.11, SD = .01) than young adults (M = 0.14, SD = .01), this difference was statistically significant. No significant interactions with age group were identified.

Table 20

Means and Standard error of errors per condition by age

Structure	Plausibility	Age	
		Younger adult	Older adults
Active	Plausible	0.03 (.01)	0.02 (.01)
	Implausible	0.17 (.02)	0.15 (.02)
Passive	Plausible	0.07 (.02)	0.04 (.01)
	Implausible	0.31 (.03)	0.26 (.03)

Table 21

GLM model of good enough processing by condition and age

	Estimate	Standard error	Z Value	
Intercept	-2.304	0.098	-23.54	***
Structure	0.715	0.196	3.65	***
Plausibility	1.979	0.196	10.11	***
Age	-0.394	0.196	-2.01	*
Structure*Plausibility	0.034	0.391	0.09	
Structure*Age	-0.183	0.391	-0.47	
Plausibility*Age	0.365	0.391	0.93	
Structure*Plausibility*Age	0.009	0.783	0.01	

*0 '***' 0.001 '***' 0.01 '**' 0.05 '*'*

6.3.3. Is there a relationship between eye movement behaviour and good enough processing?

Our third aim was to investigate the relationship between risky reading and good enough processing. A logistic regression was conducted to assess whether eye movement behaviour predicted accuracy in responses to comprehension questions. Results showed that more regressive saccades, shorter progressive saccade length, and more skips were predictors of more errors in comprehension questions (see Table 22).

To assess the relationship between risky reading, Age group, and good enough processing, Age group was added to the logistic regression (see Table 23). An interaction between Progressive Saccade Length and Age group showed that a shorter saccade length predicted more comprehension errors for younger adult participants. (See Figure 22). There were no differences for older adults.

An interaction between Regressive Saccades x Age group found that more regressive saccades were associated with more errors for older adults (See Figure 23). There were no differences found for young adults.

Additionally, a main effect of Probability of Skipping was qualified by a Probability of Skipping x Age group interaction (See Figure 24). The were no differences for older adults. In contrast, young adults who were more likely to skip words made more errors.

Table 22

GLM model of eye movement measures and Accuracy

	Estimate	Standard error	Z Value	
Intercept	-1.261	0.064	-19.80	***
Progressive saccade length	-0.206	0.023	-8.84	***
Intercept	-2.245	0.044	-51.26	***
Regressive saccade	0.036	0.003	12.38	***
Intercept	-2.160	0.042	-50.95	***
Probability of skipping	0.030	0.004	7.15	***

*0 '***' 0.001 '***' 0.01 '**' 0.05 '*'*

Table 23

GLM model of eye movement measures, Age and Accuracy

	Estimate	Standard error	Z Value	
Progressive saccade length				
Intercept	-1.382	0.067	-20.485	***
Progressive saccade length	-0.171	0.024	-6.979	***
Age	-0.714	0.135	-5.292	***
Progressive saccade length *Age	0.221	0.049	4.519	***
Regressive saccade				
Intercept	-2.285	0.046	-49.43	***

Regressive saccade	0.037	0.003	11.66	***
Age	-1.323	0.092	-14.31	***
Regressive saccade *Age	0.089	0.006	14.02	***
Probability of skipping				
Intercept	-2.159	0.043	-50.07	***
Regressive saccade	0.029	0.004	6.74	***
Age	0.120	0.086	1.39	
Skipping probability*Age	-0.055	0.009	-6.37	***

*0 '***' 0.001 '***' 0.01 '**' 0.05 '*'*

Figure 22

Young and older adults' progressive saccade length plotted by accuracy

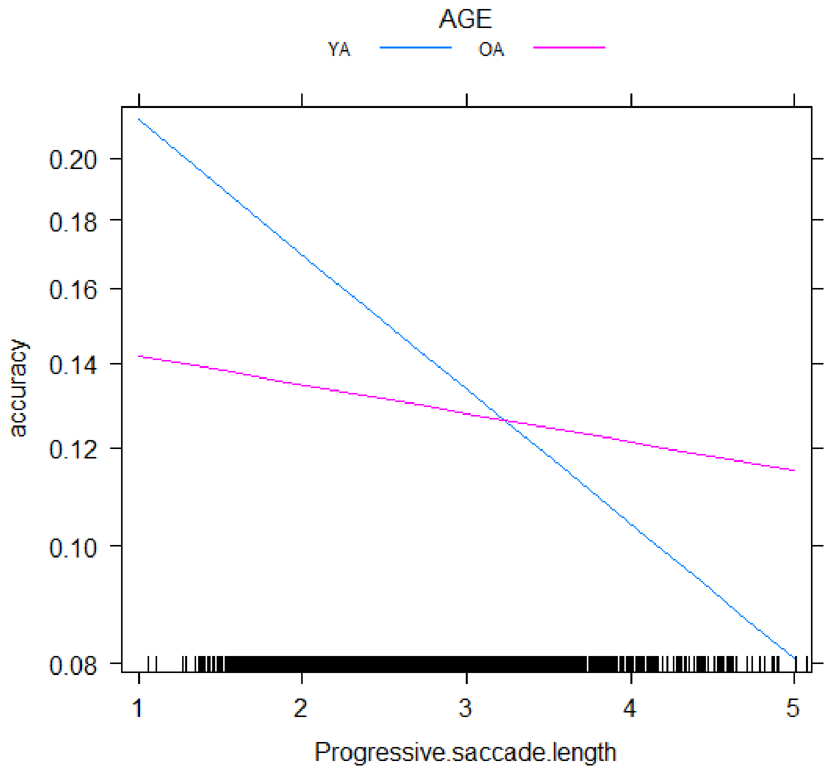


Figure 23

Young and older adults' regressive saccades plotted by accuracy

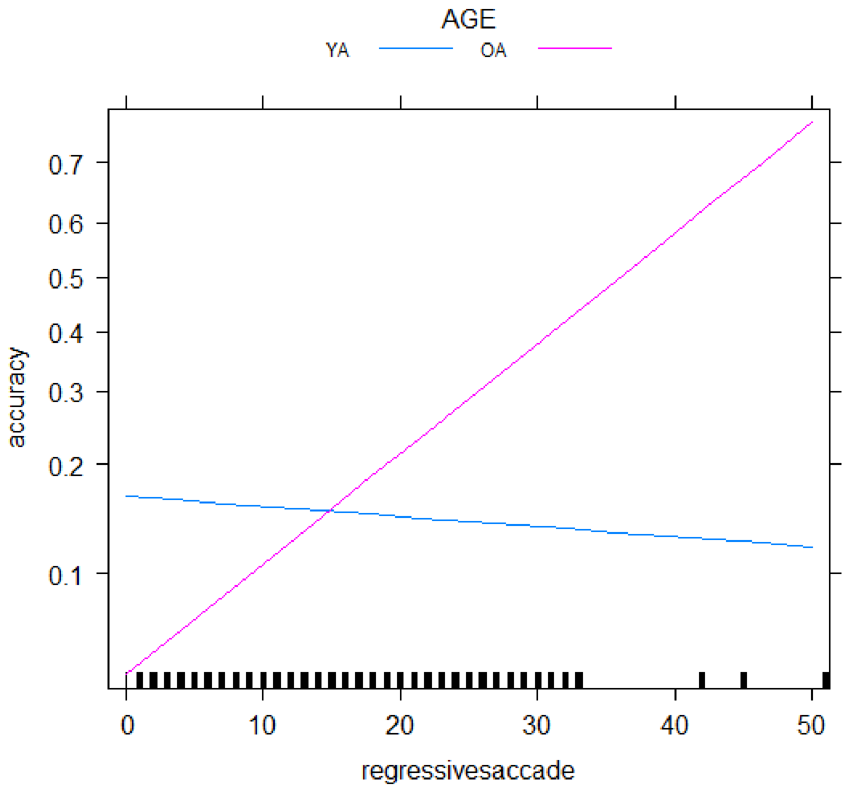
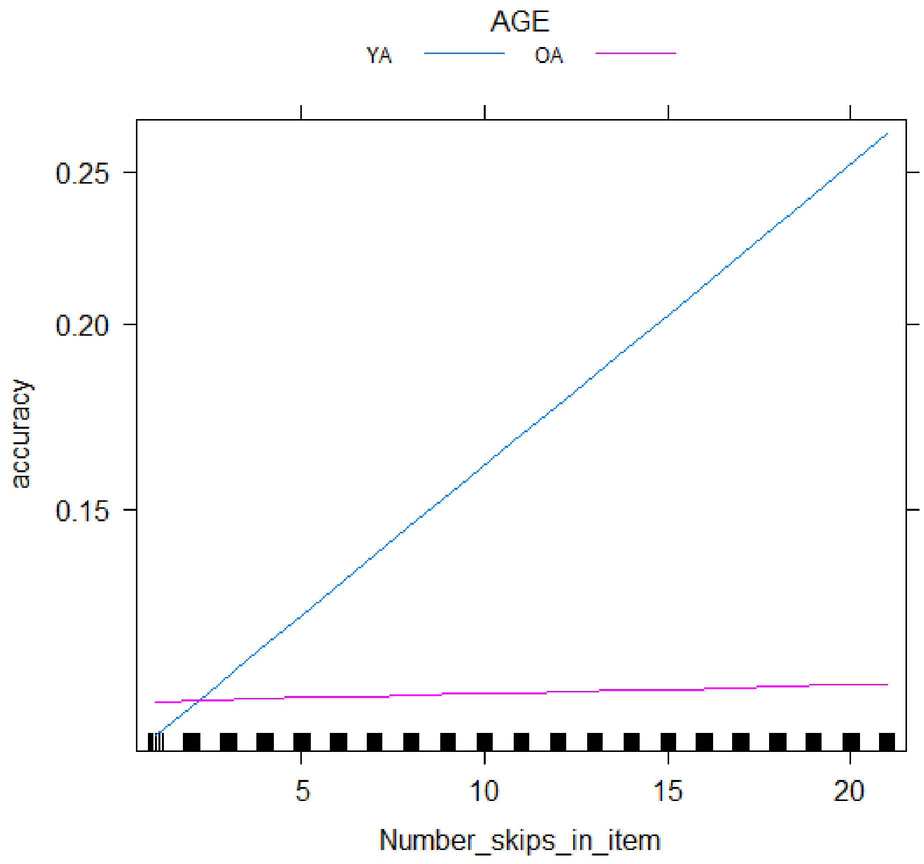


Figure 24

Young and older adults' probability of skipping plotted by accuracy



6.4. Discussion

The present study aimed to investigate whether there is a relationship between eye movements during reading and comprehension accuracy. Specifically, we examined whether young and older adults engaged in good enough processing when reading active or passive sentences describing either plausible or implausible events, whether older adults engaged in risky reading, as well as the relationship between the two. Some evidence of good enough processing was obtained, as more errors were found for passive sentences (compared to active sentences) and implausible sentences (compared to plausible sentences). Older adults made fewer comprehension errors than young adults, but age did not interact with any other factor.

In terms of eye movement behaviour, older adults showed a 'riskier' reading strategy than young adults as evidenced by longer progressive saccade length, a higher probability of skipping, and more regressive saccades. In addition, a relationship between reading behaviour and good enough processing was identified. The characteristics measured for risky reading; more regressive saccades and a high probability of skipping predicted more errors. Counterintuitively, shorter progressive saccade lengths also predicted more errors. Age differences were found; for young adults a higher probability of skipping and shorter progressive saccades were related to more errors. For older adults, more regressive saccades were related to a higher proportion of errors.

6.4.1. Good enough processing

The present findings broadly support the predictions of good enough processing theories, as passive sentences resulted in more errors than active sentences, and implausible sentences resulted in lower comprehension rates than plausible sentences. We did not fully replicate Ferreira's (2003) findings as the interaction between sentence structure and plausibility was not statistically significant, although the means did reflect this pattern. Our results contrast with recent studies that have found no evidence of good enough processing when manipulating plausibility and structure (Bader & Meng, 2018; Cutter et al., 2022; Meng & Bader, 2021). For example, Cutter et al. (2022) found no evidence of sentence canonicality affecting interpretations derived by young and older adults. However, differences in results may be due to the different task demands of each study. Cutter et al. (2022) employed a self-paced reading task and examined reading times on a subsequent sentence which was compatible with either an algorithmic or heuristic parse of the target sentence, rather than employing explicit comprehension questions to probe interpretation. Bader and Meng's (2018; Meng & Bader, 2021) 'retrieval account' suggests that any errors in passive sentences occur due to task-specific demands, specifically, the requirement to retrieve information driven by questions containing thematic probes. This account would explain the differences found in the present findings and Cutter et al. (2022), however, the current results are broadly compatible with both 'good enough' and 'retrieval based' accounts.

One effect that is more difficult for these accounts to explain is the main effect of age, with older adults engaging in less good enough processing than young adults. These findings are in contrast to those of previous studies which have examined the processing of temporary syntactic ambiguities, and showed that older adults engaged in good enough processing more so than young adults, resulting in more errors than their younger counterparts (Christianson et al., 2006). There is evidence to suggest that adults accumulate greater linguistic experience across the lifespan, often performing no worse, and in some cases outperforming young adults in vocabulary tasks (Brysbaert et al., 2016). It is interesting to note that older adults had higher vocabulary scores (and higher working memory scores) than the younger adults. In this experiment, algorithmic processing may have been preserved in older adult participants due to fluid intelligence and crystallised intelligence (Paterson et al., 2020;

Salthouse, 2009) and as a result, stronger semantic networks enabled older adults to process items correctly (Laver & Burke, 1993).

Another alternative framework is the noisy-channel inference framework (Gibson et al., 2013; Levy, 2008). The noisy-channel account explains errors in passive sentences as being due to faults in perceptual encoding and that participants may make deletions of the passive words 'was' and 'by' to create a plausible meaning. Again, this framework does not explain the age effects found in the current study as older adults would be expected to find perceptual encoding more difficult but outperformed young adults in the current study.

6.4.2. Risky reading

The present study found evidence of older adults engaging in the characteristic behaviours of risky reading, as this age group were significantly more likely to make longer progressive saccades, skip words, and make more regressive saccades than young adults. This study has furthered our understanding of risky reading behaviour in a number of respects. Firstly, the results shed light on the breadth of situations in which older adults adopt a risky reading strategy. Most previous studies reporting evidence of risky reading have used single line texts (see Zhang et al., 2022, for a full list). This study provides new evidence that it occurs during reading of multiline texts (other studies using multline text include, Payne & Stine-Morrow, 2012; Stine-morrow et al., 2010; Whitford & Titone, 2017, 2019). We also showed that risky reading occurs when processing more complex texts, such as, passive (Haeuser et al., 2021) and implausible sentences. Finally, older adults have been found to have a lower probability of skipping a word when a question is presented after every item (Wotschack & Kliegl, 2013), however the present study has identified risky reading even with these heightened task demands.

Rayner et al. (2006) suggested that older adults may engage in a risky reading strategy by using contextual information to infer the identity of the upcoming words to compensate for slower processing that occurs during healthy aging. Alternatively, a recent meta-analysis has suggested risky reading may be more indicative of non-linguistic processes, such as, mindless reading (Wotschack & Kliegl, 2013), poor saccadic control (Irving et al., 2006; Nieuwenhuis et al., 2000; Peltsch et al., 2011; Warabi et al., 1984) or that older adults are skim reading to produce gist-based interpretations (Zhang et al., 2022). While the underlying mechanisms of risky reading are currently unknown, this study has broadened our knowledge of the situations under which risky reading occurs.

6.4.3. The relationship between eye movement behaviour and good enough processing

A further aim of the current study was to assess whether there is a relationship between eye movement behaviour during reading and good enough processing. In particular, little is currently known about the consequences of a risky reading strategy for the quality of the mental representation of the content of the text. This study found a relationship between good enough processing and some of the indices of risky reading risky reading, in that more regressive saccades and more instances of skipping resulted in more errors. However, counter to expectations, a shorter progressive saccade length (rather than a longer one) was also linked to more errors.

Few published studies to date have investigated the relationship between eye movements during reading and good enough processing (Huang & Ferreira, 2021; Wonnacott et al., 2016). Recently, Huang and Ferreira (2021) conducted an eye-tracking experiment and found that longer first-pass and total reading times predicted more errors in comprehending garden-path sentences. No effect was found for go-past time. In Wonnacott et al.'s (2016) study, no relationship between eye movements and good enough processing was found. While more regressions were found to predict accuracy, this was the case for both experimental and control sentences. In the present study, the finding that more regressive saccades are associated with more errors may seem counterintuitive, however, previously studies have shown that although reanalysis may occur (Huang & Ferreira, 2021; Slattery et al., 2013; Sturt, 2007), an incorrect representation may not be successfully derived and this may be the case in this instance. Note also that more regressions are regarded as an index of risky reading, and therefore the finding that more regressions are associated with more ‘good enough’ processing supports our predictions of a positive association between risky reading and ‘good enough’ processing. Our analyses including age group as a factor demonstrated this relationship for older adult readers only, which again would be in line with our predictions. In terms of the possible mechanisms underlying this relationship, numerous studies have shown that the longer a reader holds onto a misinterpretation the more difficult it is to derive a correct interpretation; despite reanalysis, the misinterpretation can still persist. This may explain why more regressive saccades made by older adult readers predicted more errors.

In contrast to this finding for older adult results, results showed that young adult readers who skipped words more often and made shorter progressive saccades tended to also make more errors on comprehension questions. The findings in relation to skipping are in line with our hypotheses, since skipping words more frequently is also an index of risky reading behaviour. However, the findings in relation to progressive saccade length are counter to our predictions, following which longer (rather than shorter) progressive saccades should also be associated with lower comprehension accuracy. Indeed, this finding is somewhat surprising as it seems intuitive that more skipping would be associated with longer progressive saccade lengths. One possible explanation is that as young adults reanalysed the sentences, shorter progressive saccade lengths were produced. Generally, our analyses of the relationship between eye movement behaviour during reading and comprehension accuracy showed that some of the characteristics of risky reading seem to be linked to lower comprehension accuracy. This study manipulated sentence structure and plausibility, it would be interesting to further explore the relationship between risky and comprehension accuracy when other linguistic factors are manipulated.

6.4.4. Conclusion

In conclusion, we investigated age differences in ‘good enough’ processing, whether older adults engage in risky reading for more complex stimuli than those typically investigated, and the relationship between the risky reading and good enough processing. The current findings have a number of implications. Firstly, they have advanced our understanding of the different contexts in which older adults utilise a risky reading strategy. Interestingly, despite older adults’ greater linguistic and world knowledge, this age group did not engage in good enough processing more than their younger counterparts. Importantly, this study is one of few experiments which have examined the relationship between eye movements during reading and good enough processing. The results indicate that there is a relationship between online processing and comprehension accuracy. Specifically, more regressive saccades, shorter progressive saccade length, and more skipping were associated with more good enough processing. Furthermore, age differences in this relationship were identified. Young adults who were more likely to skip words and had shorter progressive saccades tended to make more errors on comprehension questions. In contrast, for older adult readers, more regressive saccades predicted lower accuracy. These findings suggest that certain indices of risky reading behaviour may have consequences for offline comprehension in relation to 'good enough' interpretations - future work could consider other linguistic consequences of risky reading.

In Chapter 7, Experiment 4 explored young and older adults' susceptibility to shallow processing at the discourse-level using semantic ambiguity. Participants were presented with sentences that contained doubly quantified sentences (e.g., *Every kid climbed a tree*), these are ambiguous regarding the number of entities that are represented by the indefinite phrase (*a tree*). A context sentence biased participants to a plural or singular interpretation. The use of eye-tracking enabled the investigation of on-line processing of subsequent anaphoric reference to the indefinite phrase (e.g., *The tree was* vs. *The trees were*). Similarly to Experiment 1 and 2, working memory was also assessed.

Chapter 7: Quantifier scope ambiguity

7.1. Introduction

When a sentence contains two or more quantifiers, ambiguity can arise regarding the number of entities to be represented. For instance, when reading a sentence such as *Every kid climbed a tree* (containing the quantifiers *every* and *a*), while it is clear that there is more than one kid, the sentence is ambiguous as to whether there is also more than one tree. How this ambiguity is interpreted is said to depend on which quantifier takes 'wide scope' (Kurtzman & MacDonald, 1993).

If *every* takes wide scope over *a*, the sentence is interpreted as meaning that each kid climbed a different tree, so there may be many trees. In this instance, a plural continuation such as (1a) should be easier to process than a singular continuation such as (1b). In contrast, if *a* takes wide scope over *every*, the sentence is interpreted as meaning every kid climbed the same tree, so there is only one tree. In this case, a singular continuation (e.g., 1b) would be easier to process.

1a) Every kid climbed a tree. The trees were…

1b) Every kid climbed a tree. The tree was…

Several factors have been thought to influence which quantifier takes wide scope, with most research focusing on structural (such as the order in which the quantifiers appear in the sentence) and semantic aspects (such as properties of the quantifier itself). In contrast, the influence of pragmatic factors, as well characteristics of the reader which might influence their ability to process ambiguity, have been relatively underexplored. In the present paper, we investigate the influence of world knowledge, aging, and working memory capacity, as well as examining how these factors may interact with structural and semantic cues.

Several accounts have aimed to explain the mechanisms underlying the processing of quantifier scope ambiguity. Under one account (Lakoff, 1971), structural factors, specifically, the linear order of quantifiers within a sentence, are the principal determinant; with the first quantifier taking wide scope. Alternatively, Ioup (1975) advocated a quantifier hierarchy (see 2), where semantic factors mean that certain (stronger) quantifiers have a tendency to take wide scope over other (weaker) quantifiers.

2) each > every > a > all > most > many > several > some > a few

These classical theoretical approaches focus solely on the influence of structural or semantic factors, and their predictions have been tested in previous empirical research examining how quantifier scope ambiguities are processed on-line (e.g., Filik et al., 2004, Paterson et al. 2008). However, it is clear from research in other domains of psycholinguistics that context can play a major role in the moment-to-moment processes underlying ambiguity resolution - from resolving lexical ambiguity (Duffy et al., 1988), to higher level processes involved in interpreting statements that are ambiguous between a literal and figurative interpretation (Turcan et al., 2020). Therefore, it seems likely that contextual bias plays an important role in the processing of quantifier scope ambiguities as well.

An alternative theoretical position which allows several factors to interact in determining which interpretation of a quantifier scope ambiguity is adopted is the constraint-satisfaction approach (Kurtzman & MacDonald, 1993). Notably, the factors that may contribute to this process during on-line reading are yet to be determined. Thus, the aim of the present study is to gain a fuller picture of the influence of a range of factors. As well as the influence of context, we examine potentially important individual differences and cognitive factors, specifically, age and working memory. The reason for focusing on these particular individual differences is that there is some preliminary evidence to suggest that these may play a role (see Kemtes & Kemper, 1999); however, their influence on on-line processing is currently unclear.

The foundation for the present study are the experiments reported in Filik et al. (2004) and Paterson et al. (2008), who conducted eye-tracking while reading studies to investigate the influence of structural and semantic factors. In their studies, participants read ambiguous doubly quantified sentences that were followed by singular or plural anaphoric reference to the ambiguous phrase. Filik et al. (2004) used dative (e.g., 3a and 3b) and double object sentences (e.g., 3c and 3d) and manipulated the order in which the quantifiers *every* and *a* appeared in the sentence. Paterson et al. (2008) conducted a similar study using *each* instead of *every*.

3a) The celebrity gave an in-depth interview to every reporter from the newspaper, but the interview(s) was/were not very interesting. (dative / *a–every*)

3b) The celebrity gave every in-depth interview to a reporter from the newspaper, but the reporter(s) was/were not very interested. (dative / *every–a*)

3c) The celebrity gave a reporter from the newspaper every in-depth interview, but the reporter(s) was/were not very interested. (double object / *a–every*)

3d) The celebrity gave every reporter from the newspaper an in-depth interview, but the interview(s) was/were not very interesting. (double object / *every–a*)

Despite interactions found on the region of the sentence containing the two quantified phrases, longer reading times were observed for plural than singular continuations (e.g., *the interviews/reporters were* vs. *the interview/reporter was*) irrespective of linear order or quantifier hierarchy. These findings suggest that structural and semantic factors such as linear order and properties of the individual quantifiers themselves may have little influence on the on-line processing of subsequent noun phrase anaphors. The authors argued from this that readers may have a default preference for a singular interpretation, or that they may instead adopt a superficial surface matching strategy based on a match between the morphological features of the anaphor and its antecedent (which is always quantified by *a* and therefore always superficially singular).

From these eye-tracking studies, it appears that linguistic factors have a limited influence on the on-line processing of anaphoric reference to quantified entities during normal reading. Some preliminary evidence that non-linguistic factors, specifically, age and working memory, may be important was provided by Kemtes and Kemper (1999). In two self-paced reading experiments, Kemtes and Kemper presented ambiguous doubly quantified sentences (e.g., 4 and 5), plus unambiguous controls, to young and older adults. In Experiment 1, participants read the sentence containing the two quantifiers, and then judged whether the continuation sentence made sense. Following a lack of strong observable preferences in Experiment 1, in Experiment 2, the same sentences were presented with the addition of a context sentence (such as *The kids were in the park*, or *The trees were in the park*) that was designed to encourage participants to process the texts more deeply.

4a) A kid climbed every tree. The kid was full of energy.

4b) A kid climbed every tree. The kids were full of energy.

5a) Every kid climbed a tree. The tree was full of apples.

5b) Every kid climbed a tree. The trees were full of apples.

The results of Experiment 2 showed that when the quantifier order was *a-every* (as in

4a and 4b), both young and older adults preferred the singular noun phrase (NP) anaphor (*The kid was…*). When the order was *every-a* (as in 5a and 5b), young adults preferred plural continuations (*The trees were…*), however, older adults still had a singular preference (*The tree was…*). Context did not affect these preferences. The authors concluded from this that young adults used structural principles when resolving quantifier scope ambiguities. In contrast, they proposed that older adults (in particular, those with a low working memory span) may rely on a pragmatic principle which Kemtes and Kemper termed the "principle of single reference". Following this principle, a single entity is always assumed whenever the quantifier *a* is encountered. In general, across both experiments, older adults had longer reading times for the quantifier sentence than young adults. However, no age differences in reading times were observed in relation to context or quantifier order. Kemtes and Kemper concluded from this that working memory affects post comprehension decision processes, but not reading times. However, it is important to note that Kemtes and Kemper used self-paced reading in combination with an explicit judgement task, which may not reflect processes involved during natural reading.

7.1.1. Present study and predictions

In the present study, young and older adult participants had their eye movements monitored while they read doubly quantified sentences that, on the basis of their world knowledge, would be expected to bias the reader towards either a singular or plural referent for the indefinite phrase. For example, in a context such as *Every student in the class is listed on a register*, our world knowledge would suggest that each student is listed on the same register, implying a singular referent. In contrast, a context such as *Every school pupil in the country has their attendance marked on a register* implies that there is more than one register. The quantifier sentence was then followed by a continuation sentence containing either a singular or plural anaphoric reference (e.g., *This register* vs. *These registers*) to the indefinite phrase (see Table 24). Following the eye-tracking experiment, participants' working memory capacity was assessed.

Table 24

Example material in all conditions with analysis regions.

Condition		Example material
Context	Reference	
Singular	Singular	Every student in the class is listed on a register.\ This register(critical region)\ is routinely kept (post-critical region)\ in a very safe place. Pupils receive a certificate for good attendance at the end of the year.
Singular	Plural	Every student in the class is listed on a register.\ These registers(critical region)\ are routinely kept (post-critical region)\ in a very safe place. Pupils receive a certificate for good attendance at the end of the year.
Plural	Singular	Every school pupil in the country has their attendance marked on a register.\ This register (critical region)\ is routinely kept(post-critical region)\ in a very safe place. Pupils receive a certificate for good attendance at the end of the year.
Plural	Plural	Every school pupil in the country has their attendance marked on a register. \ These registers (critical region)\ are routinely kept (post-critical region)\ in a very safe place. Pupils receive a certificate for good attendance at the end of the year.

In terms of predictions, although we are not manipulating the linear order of the quantifiers in the sentence, following the linear order principle (Lakoff, 1971), if readers assign wide scope to the first quantifier, then we would expect them to assign wide scope to *every* (since all sentences begin with *every*). This would lead to the mental representation of multiple entities (e.g., multiple registers), and should result in shorter reading times for plural than singular NP anaphors (e.g., *These registers* vs. *This register*). This pattern of effects would also be predicted by Ioup's (1975) quantifier hierarchy, which states that *every* should take wide scope over *a*. Thus, if structural factors are driving processing, we would expect shorter reading times for plural than singular NP anaphors. In contrast, if pragmatic factors are driving processing, then we would instead expect greater processing difficulty when the number of entities referred to by the NP anaphor does not match the pragmatic bias introduced in the context. Following a model in which a number of principles can interact to simultaneously influence processing (e.g. Kurtzman & Macdonald, 1993), we might expect stronger effects for plural biasing contexts, since in these contexts the pragmatic bias towards a plural referent will support the linear order and quantifier hierarchy biases (which would also predict a plural referent) whereas in the singular biasing context condition, the pragmatic constraint would be in opposition to these other factors.

Based on previous eye-tracking findings, if it is the case that when reading the NP anaphor, readers engage in a superficial 'surface matching' strategy (Filik et al., 2004; Paterson et al., 2008), then we would expect shorter reading times on singular than plural NP anaphors, regardless of contextual bias (for young adults at least). In relation to the judgement task findings of Kemtes and Kemper (1999), we might expect context to have no influence on either age group. Instead, we might expect older adults, especially those with low working memory capacity, to display shorter reading times for singular than plural NP anaphors if they follow the pragmatic principle of single reference. In contrast, for younger adults, who would be expected to follow structural principles (such as those outlined above which would predict wide scope to be assigned to *every*), we would expect the opposite effect, that is, shorter reading times for plural than singular references. However, note that these latter predictions are based on judgement data rather than reading times and so the processes involved in this task versus online processing may differ.

7.2. Method

7.2.1. Participants

Forty-eight young (39 female, M_{age} = 20 years old, age range = 18-26) and 48 older adults (30 female, M_{age} = 70 years old, age range = 65-77) took part. All participants were native English-speakers and were recruited from the University of Nottingham and surrounding area. Older adults received an honorarium for their participation. Young adults were provided with course credit or an honorarium for their involvement in the study. All participants had normal or corrected vision and did not report any history of dyslexia. The study was approved by the University of Nottingham's School of Psychology ethics board.

Visual acuity and contrast sensitivity tests were conducted to ensure that older participants did not suffer any visual or cognitive impairment beyond what would be expected with normal aging (see Table 10). Specifically, visual acuity was assessed using a SLOAN ETDRS chart, and a SLOAN eye chart was used to assess contrast-sensitivity. Each of these tests were conducted at the recommended distance of 40cm. Cognitive abilities were examined using the Montreal Cognitive Assessment Test (MoCA) (Nasreddine et al., 2005). Working memory was examined using the Wechsler Adult Intelligence Test (WAIS-IV) (Wechsler, 2008), and was included in the analyses for the main experiment.

Young adults had better visual acuity than older adults, $t(94) = 6.43$, $p < .001$, d = 1.31 [95% CI -9.04 to -4.77]. Young adults also had better contrast sensitivity than older adults, $t(94) = 7.00$, $p < .001$, $d = 1.43$ [95% CI -19.15 to -10.68]. Older adults typically perform worse on vision tests compared to their younger counterparts (Elliot et al., 1995). Importantly, all participants had visual acuity of up to 20/40, which is within clinically acceptable limits, and all older adults scored within the healthy range on the MoCA (≥ 26).

7.2.2. Design and Materials

The experiment had a 2 (*Context*: singular-biasing vs. plural-biasing) x 2 (*Reference*: singular vs. plural) x 2 (*Age*: young vs. older adults) mixed-design, with Context and Reference as within-subjects factors and Age as a between-subjects factor.

Participants read 96 trials in total. Twenty-four of these were the experimental items for the current study (see Appendix - item E). Of the remaining 72 trials, 32 were experimental items from another Experiment 2 (Chapter 5). The final 40 items consisted of a variety of syntactic complexities to avoid task effects and prevent readers from adopting strategies (see Appendix - item F).

Each experimental trial consisted of three sentences: a doubly quantified sentence, followed by a target continuation sentence, and finally a post-target sentence (see Table 1). The doubly quantified sentence contained the quantifiers *every* and *a*. It described a scenario which biased either a singular or a plural referent of the phrase quantified by *a*, based on the reader's knowledge of the world. The subsequent target sentence contained either a singular or plural anaphoric reference to the indefinite phrase. The final sentence was included so that the target sentence was not at the end of the trial. Each item was presented across three lines, with two line spaces between each line of text. A binary yes/no comprehension question followed 25% of experimental items - these questions assessed engagement in the task (see Appendix – E). There were four counterbalanced presentation lists resulting in participants seeing six items in each of the four conditions. The presentation list viewed by each participant was presented in a random order.

Norming task: The materials were normed to ensure that the items biased individuals as intended. Sixteen native English speakers read the experimental items and selected from a choice of accompanying diagrams whether an item had a singular or plural interpretation. The average accuracy score was 0.88 (*SD* = 0.33). Correct interpretations for plural items were slightly higher (*M* = 90%) than singular items (*M* = 86%), but this was not statistically significant X^2 (1, *n* = 768) = 2.05, *p* = .15, *phi* = -.06.

7.2.3. Apparatus

An SR Research EyeLink 1000 desktop mounted eye-tracker was used to record gaze location every millisecond. Only the right eye was tracked although viewing was binocular. Sentences were displayed in Courier New 14-point font as black text on a white background on a 17-inch CRT monitor at a viewing distance of 58cm. Three characters subtended approximately one degree of visual angle. A chin and forehead rest were used to minimise movements.

7.2.4. Procedure

Initially, participants completed the demographic questions and vision tests. In addition, older adults completed the cognitive ability assessments. Next, the eye-tracking task was completed. All participants were instructed to read the short texts as they would normally, for comprehension, and that questions would appear after some of them. Prior to the presentation of the trials, a nine-point calibration procedure was conducted (average error <.5 degrees of visual angle), and the accuracy of the calibration was checked before each trial. Following this, participants fixated on a box located on the top left-hand side of the screen, and the sentence was automatically presented with the first letter of the item replacing the box. Participants were instructed to press a button on a PC game pad once they had finished reading the item. The text then disappeared and was replaced by a question, or if there was no question, then a calibration target appeared. To answer the questions, participants pressed the left-hand shoulder button on the game pad for *no* and right button for *yes*. Once the question had been answered, a calibration target appeared. A break was offered every third of the way through the trials and it took approximately 45 minutes for younger adults to complete and one hour for older adults. After the eye-tracking session, the WAIS-IV digit-span forward, digit-span backward and digit-span sequential tests were administered for a comprehensive working memory score.

7.2.5. Analyses

We report four measures of eye movement behaviour (following previous work in our lab examining pragmatic processing in young and older adults, Howman & Filik, 2020). The following two eye movement measures are reported as indices of early processing difficulty: *First fixation duration* - defined as the duration of the first fixation on a region of interest, and *first-pass reading times* - the sum of all fixation durations from the eyes first entering the specific region to first leaving it. *Regression path reading time* was considered an index of late processing difficulty and is the time from the eyes first fixating on a given region plus all time spent fixating to the left of the region before fixating on a subsequent region to the right for the first time. Regression path reading times are indicative of time spent resolving issues encountered, which may involve re-reading previous text before moving on from the region where processing difficulty is experienced. The sum of all fixations in a region constitutes *total reading time*, and indicates overall difficulty in a given region.

Analyses were conducted for the critical region (NP anaphor, e.g., *This register/These registers*) and the post-critical region (e.g., *is routinely kept/are routinely kept*) (see Table 1). Short contiguous fixations were combined in an automatic procedure. Fixations under 40 ms that were not within three characters of another fixation were deleted, as were fixations over 1200 ms. Fixations under 80ms were integrated into larger adjacent fixations within one character. Trials where participants failed to read the sentence or where track-loss occurred were eliminated. That is, trials where two or more adjacent regions had zero first-pass reading times were removed, accounting for 1.05% of the data.

Prior to each individual analysis, all trials with zeros for each continuous reading time measure in each region were removed. Furthermore, a conservative criterion value of 2.5 standard deviations was set per participant, per condition for each dependent variable and any values above this were removed. In the critical region, this procedure accounted for 5.23% of first fixation durations, 5.27% of first-pass reading times, 5.50% of regression path reading times, and 1.00% of total reading times. In the post-critical region, 1.41% of data in first fixation durations, 1.55% of first-pass reading times, 1.55% of regression path reading times, and 0.59% of total reading times data were removed.

7.3. Results and Discussion

A generalised linear mixed effects model (GLMM; Baayen et al., 2008) was conducted using the lme4 package (Baayen et al., 2008; Bates et al., 2015) to analyse the data in R (Version 3.5.1; R Core Team, 2019). The gamma family and identity link were utilised to accommodate the skew of the data and removes the need to log-transform data which enables interpretation without back transforming (Lo & Andrews, 2015).

Context, reference, age, and working memory were fixed factors and items and subjects were random factors. At first, a full random structure of participants and items was performed to avoid being anti-conservative (Barr et al., 2013). If a model failed to converge, the number of iterations was increased, the optimizer “bobyqa” was included and then the model was trimmed - removing correlations, interactions, and slopes until convergence was successful. Only results from successful model convergence are reported (see Appendix – Item D for the syntax of the final models). *P* values were automatically calculated from the lmerTest package (Kuznetsova et al., 2017). All categorical factors were contrasted using the MASS package (Venables & Ripley, 2002) and function contr.sdif (0.5/-0.5) meaning that the intercept corresponds to the grand mean and the fixed effects to the main effect of the fixed factors. Simple effects were conducted for interactions identified in the GLMM (see Appendix – Item H). Descriptive statistics can be found in Tables 25 and 26.

Table 25

Means (M) and standard error (SE) for eye movement measures by age group for critical and post-critical regions

	Critical region (noun phrase anaphor)		Post-critical region	
Age group	YA	OA	YA	OA
First fixation durations (msec)				
Plural context, singular reference	205 (7)	214 (6)	227 (5)	249 (5)
Plural context, plural reference	209 (6)	220 (5)	225 (5)	241 (4)
Singular context, singular reference	207 (7)	226 (6)	217 (5)	237 (4)
Singular context, plural reference	230 (8)	231 (6)	223 (5)	238 (5)
First-pass reading times (msec)				
Plural context, singular reference	320 (14)	345 (13)	420 (17)	472 (16)
Plural context, plural reference	380 (15)	378 (13)	397 (14)	426 (12)
Singular context, singular reference	345 (14)	352 (12)	382 (13)	458 (14)
Singular context, plural reference	383 (15)	403 (14)	426 (17)	433 (16)

Regression path reading times (msec)				
Plural context, singular reference	446 (22)	442 (18)	572 (30)	578 (21)
Plural context, plural reference	557 (28)	507 (21)	500 (24)	488 (17)
Singular context, singular reference	493 (31)	540 (26)	531 (27)	529 (17)
Singular context, plural reference	580 (35)	570 (28)	519 (24)	544 (28)
Total reading times (msec)				
Plural context, singular reference	516 (19)	597 (22)	650 (27)	789 (31)
Plural context, plural reference	542 (21)	583 (20)	549 (19)	643 (24)
Singular context, singular reference	509 (19)	571 (19)	571 (21)	678 (24)
Singular context, plural reference	575 (20)	664 (23)	595 (25)	661 (26)

Table 26

Means (M) and standard error (SE) for comprehension accuracy by age group for critical and post-critical regions

Condition	Accuracy (% correct)	
	Young adults	Older adults
Plural context, singular reference	93 (3)	93 (3)
Plural context, plural reference	95 (2)	91 (3)
Singular context, singular reference	96 (2)	95 (2)
Singular context, plural reference	94 (2)	95 (2)

All models included context, reference, and age, as these were theoretically fundamental to our hypotheses. For the exploratory covariate, working memory, separate models were run, which included the fixed factors from the base model and a composite working memory score which interacted with the other fixed factors. Finally, maximum likelihood (ML) estimates were used to compare models. Working memory was not found to be a better fit in comparison to the base model for any of the measures. Only results of models of best fit are presented (see Table 27 and 28).

7.3.1. Comprehension accuracy

Participants who scored less than 75% correct in the comprehension questions were removed from the analysis. This resulted in data being removed from two participants (one young and one older adult). For the remaining participants, overall comprehension was 95% (*SE* = 1%) for the young adults and 94% (*SE* = 1%) for older adults (see also Table 26). There were no significant effects of Context, Reference, or Age on comprehension accuracy (see Table 27).

7.3.2. Critical region (e.g., This register/These registers)

A main effect of context showed longer regression path reading times in singular-biasing than plural-biasing conditions. This likely reflects the conflict between pragmatic versus structural and semantic factors in singular-biasing conditions, compared to plural-biasing conditions (in which pragmatic, structural, and semantic factors are aligned in support of a plural interpretation). A main effect of reference showed that plural NP anaphors resulted in longer first-pass, regression path, and total reading times compared to singular NP anaphors. This is in line with the results of previous eye-tracking studies (e.g., Filik et al., 2004; Paterson et al., 2008).

Table 27

Binomial mixed effect of accuracy, condition and age group

Accuracy Measure	Estimate	SE	t/z value	
Intercept	3.116	0.487	6.39	***
Bias	0.544	0.861	0.63	
Reference	0.024	0.350	0.07	
Age	0.267	0.353	0.76	
Bias x Reference	0.070	0.695	0.10	
Bias x Age	0.027	0.744	0.04	
Reference x Age	0.248	0.689	0.36	
Bias x Reference x Age	0.408	1.372	0.30	

Table 28

GLME results for critical and post-critical regions

	Critical region				Post-critical region			
	Estimate	SE	t/z value		Estimate	SE	t/z value	
First fixation durations[1]								
Intercept	232.399	5.508	42.19	***	238.401	5.004	47.64	***
Context	9.829	6.354	1.55		-4.73	6.061	-0.78	
Reference	4.547	3.177	1.43		1.286	2.595	0.50	
Age	8.438	8.384	1.01		15.667	7.623	2.06	*
Context*Reference	17.988	6.634	2.71	**	4.819	4.898	0.98	
Context*Age	-3.496	6.109	-0.57		-1.105	7.482	-0.15	
Ref*Age	-5.065	6.135	-0.83		-3.779	5.09	-0.74	
Context*Reference*Age	-14.356	10.794	-1.33		9.691	7.68	1.26	

First-pass reading times[1]								
Intercept	390.149	9.880	39.49	***	455.859	11.423	39.91	***
Context	17.644	6.187	2.85	**	1.458	9.451	0.15	
Reference	39.050	9.099	4.29	***	-8.987	6.320	-1.42	
Age	8.535	13.001	0.66		36.813	12.580	2.93	**
Context*Reference	0.961	9.496	0.10		28.861	9.496	3.04	***
Context*Age	-6.910	8.837	-0.78		5.395	9.489	0.57	
Ref*Age	-10.729	9.845	-1.09		-25.266	9.019	-2.80	*
Context*Reference*Age	31.178	13.517	2.31	***	-11.289	12.427	-0.91	
Regression-path reading times								
Intercept	554.916	10.886	50.98	***	560.514	12.271	45.68	***
Context	53.107	10.698	4.96	***	-2.093	7.403	-0.28	
Reference	59.598	12.306	4.84	***	-36.023	8.074	-4.46	***

[1] The based models in first-fixation and first pass reading times were not deemed the better model compared to when length and frequency were included. These results are reported for transparency but are not described in the results section.

Age	-4.276	9.529	-0.45		5.264	15.827	0.33	
Context*Reference	-0.668	10.946	-0.06		39.92	10.51	3.80	***
Context*Age	24.709	11.010	2.24	*	-4.213	17.63	-0.24	
Ref*Age	-48.447	11.715	-4.14	***	4.359	9.638	0.45	
Context*Reference*Age	-4.125	11.216	-0.37		19.69	10.597	1.86	
Total reading times								
Intercept	589.182	10.781	54.65	***	683.543	9.143	74.76	***
Context	19.743	10.888	1.81		-30.520	9.491	-3.22	*
Reference	31.780	8.942	3.55	***	-58.811	8.636	-6.81	***
Age	55.259	11.403	4.85	***	78.311	9.155	8.55	***
Context*Reference	54.487	10.093	5.40	***	53.263	8.290	6.43	***
Context*Age	15.285	10.648	1.44		-26.960	14.156	-1.91	
Ref*Age	-11.150	11.079	-1.01		-17.226	12.254	-1.41	
Context*Reference*Age	59.611	13.568	4.39	**	25.796	9.336	2.76	***

Significant codes: <.001’***’, <.01’**’, <.05’*’

Table 29

GLME results for first-fixation and first-pass reading times in the critical with length and frequency included

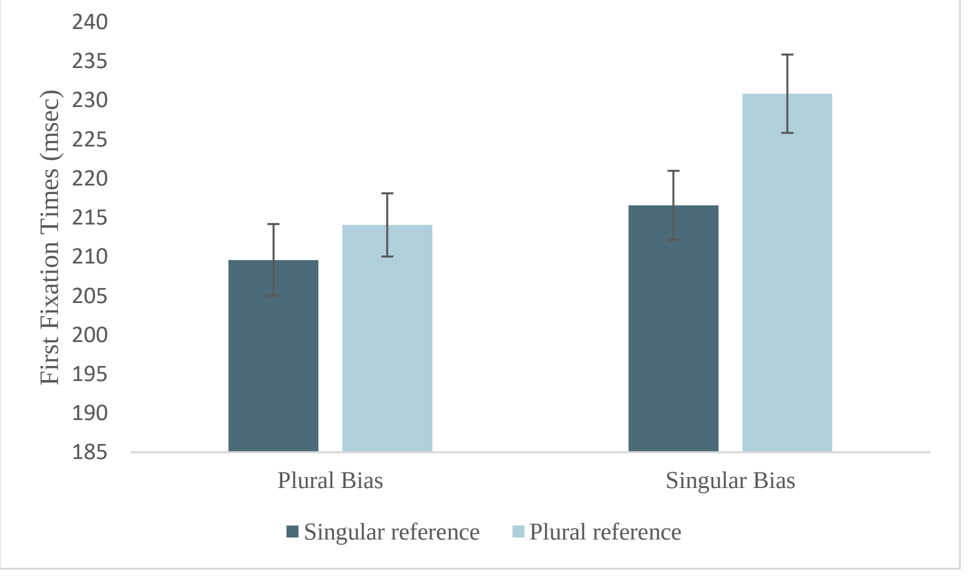
	First fixation durations				First-pass reading times			
	Estimate	SE	t/z value		Estimate	SE	t/z value	
Intercept	224.966	11.368	19.79	***	418.386	12.228	34.21	***
Context	10.919	7.647	1.43		1.197	10.226	0.12	
Reference	4.481	5.139	0.87		35.666	10.349	3.45	***
Age	8.235	9.654	0.85		12.841	15.217	0.84	
Frequency	1.823	2.654	0.69		-6.459	3.823	-1.69	
Context*Reference	19.063	5.945	3.21	**	2.773	11.708	0.24	
Context*Age	-5.048	7.996	-0.63		-5.088	11.144	-0.46	
Ref*Age	-6.570	7.672	-0.86		-4.870	11.218	-0.43	
Context*Reference*Age	14.087	9.741	-1.45		20.051	12.029	1.67	
Intercept	228.3459	8.5053	26.85	***	274.979	14.119	19.48	***

Context	9.9328	7.156	1.39		1.6060	8.535	0.19	
Reference	3.0434	5.0041	0.61		12.488	8.182	1.53	
Age	8.5615	7.4872	1.14		8.180	10.278	0.80	
Length	0.3668	0.7987	0.46		9.966	1.573	6.33	***
Context*Reference	19.1033	5.2844	3.62	***	7.875	10.030	0.79	
Context*Age	-5.0008	8.333	-0.60		-7.563	9.860	-0.77	
Ref*Age	-5.5676	6.454	-0.86		-7.788	9.072	-0.86	
Context*Reference*Age	-14.0472	8.680	-1.62		16.093	11.327	1.42	

Interactions between context and reference were also observed in first fixation duration[2] (see Figure 25) and total reading times, indicating that pragmatic factors can have an early influence on the processing of anaphoric reference to quantified phrases - greater disruption occurred when there was a mismatch in number between the expectation set up by the singular context and the subsequent plural (619 msec) and singular NP anaphor (540 msec). There was no difference in reading times for plural bias sentences followed by a singular (556 msec) or plural (563 msec) NP anaphor. These results demonstrate an influence of world knowledge on processing of subsequent NP anaphors at the critical region, with greater difficulty encountered when there is a mismatch between a singular contextual bias and the number of entities that are referred to by the NP anaphor.

Figure 25

First fixation times at the critical region



[2] To ensure that the effects encountered for reference were not due to length or frequency effects, length and frequency of the reference were added as a covariate into first fixation and first-pass reading times measures as this is typically where length and frequency effects would occur (Kliegl et al., 2004; Pollatsek, Juhasz et al., 2008; Rayner et al., 1998). There were no effects of frequency or length on first-fixation times. There was no effect of frequency on first-pass reading times but there was a length effect (see Table 29).

7.3.3. Age differences

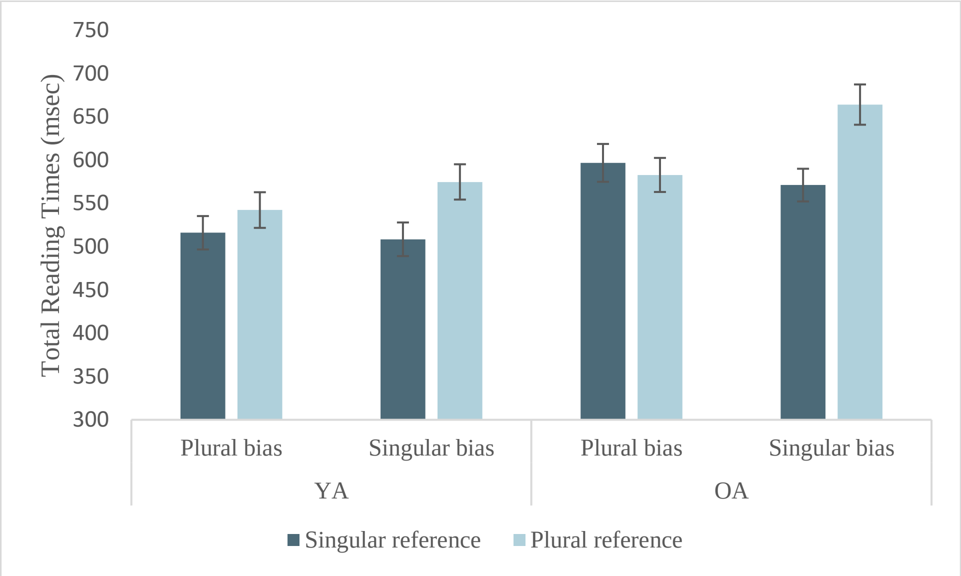
A main effect found older adults (603 msec) had longer total reading times than younger adults (535 msec). This is commensurate of research that has previously found generally longer reading times for older adults (e.g. Howman & Filik, 2020; Kemper & McDowd, 2006; Kemper & Liu, 2007; Kemper et al., 2004; Kliegl et al., 2004; Liu et al., 2017; McGowan et al., 2014; 2015; Rayner et al., 2006).

There was a Context x Age interaction in regression path reading times. Older adults had shorter regression path reading times for plural (475 msec) than singular contexts (554 msec). There was no difference in context for young adults (plural = 502 msec; singular contexts = 535 msec). There was a Reference x Age interaction in regression path reading times. Young adults had short regression path reading times for singular NP anaphors (470 msec) than plural NP anaphors (568 msec), which is in line with previous research showing a singular preference in younger adult readers (Filik et al., 2004; Paterson et al., 2008). There was no NP anaphors preference for older adults (plural = 538 msec; singular = 492 msec).

There was also a Context x Reference x Age interaction in first-pass reading times and total reading times. For first-pass reading times no age differences between condition were identified. The interaction for total reading times showed that compared to young adults, older adults had longer reading times for singular-biasing context sentences followed by singular NP anaphors, singular-biasing context sentences followed by plural NP anaphors, and plural-biasing context sentences followed by a singular NP anaphors (see Figure 26). However, there was no significant difference between younger adults and older adults for plural-biasing context sentences followed by a plural NP anaphor. This suggests that in later measures of reading time, older adults found processing as easy as younger adults when all of the constraints supported the same interpretation (i.e., plural-biasing contexts followed by plural anaphors), but experienced more difficulty when there was conflict between the different constraints (i.e., all other conditions).

Figure 26

Total reading time at the critical region



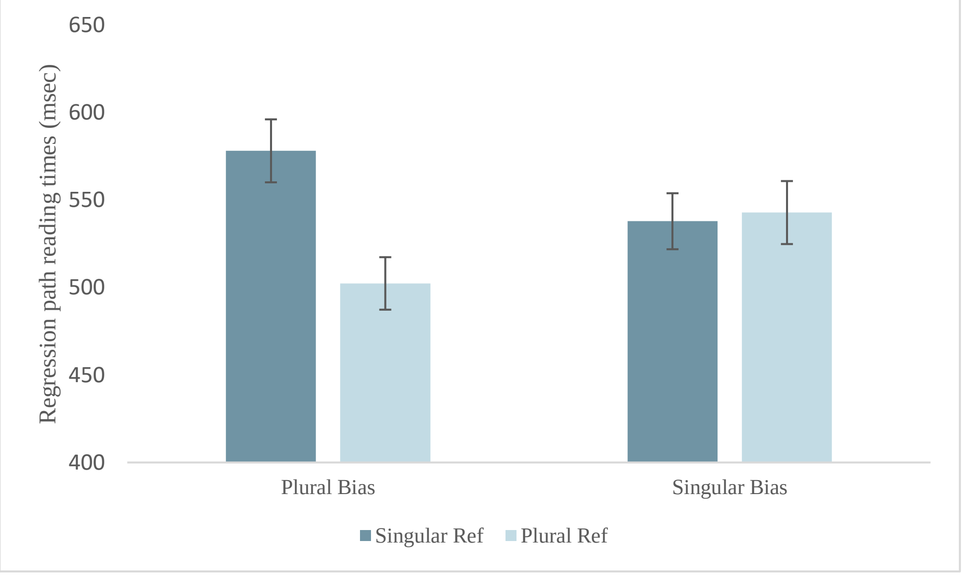
7.3.4. Post-critical region (is/are routinely kept)

There was a main effect of Context, showing longer total reading times following plural- than singular-biasing contexts. There was also a main effect of Reference, showing longer regression path and total reading times following singular than plural NP anaphors. However, these main effects were qualified by Context x Reference interactions.

Specifically, a series of Context x Reference interactions indicated that context continued to exert an influence on processing in the post-critical region with more disruption when there was a mismatch between plural bias introduced in the context and the number of entities referred to by the subsequent NP anaphor. That is, there were longer first-pass reading times when plural-biasing context sentences were followed by a singular NP anaphor (448 msec) rather than a plural NP anaphor (418 msec). There was no difference for singular contexts followed by singular (426 msec) or plural NP anaphors (438 msec). There were longer regression path (see Figure 27) and total reading times for singular NP anaphors preceded by plural-biasing (total reading times = 719 msec) than plural NP anaphors (total reading times = 596 msec). There was no difference for singular contexts followed by singular (total reading times = 625 msec) or plural NP anaphors (total reading times = 629 msec). These findings show that readers experienced processing difficulty when the number of entities referred to in the subsequent NP anaphor did not match the bias of the context sentence.

Figure 27

Regression path reading times at the post-critical region



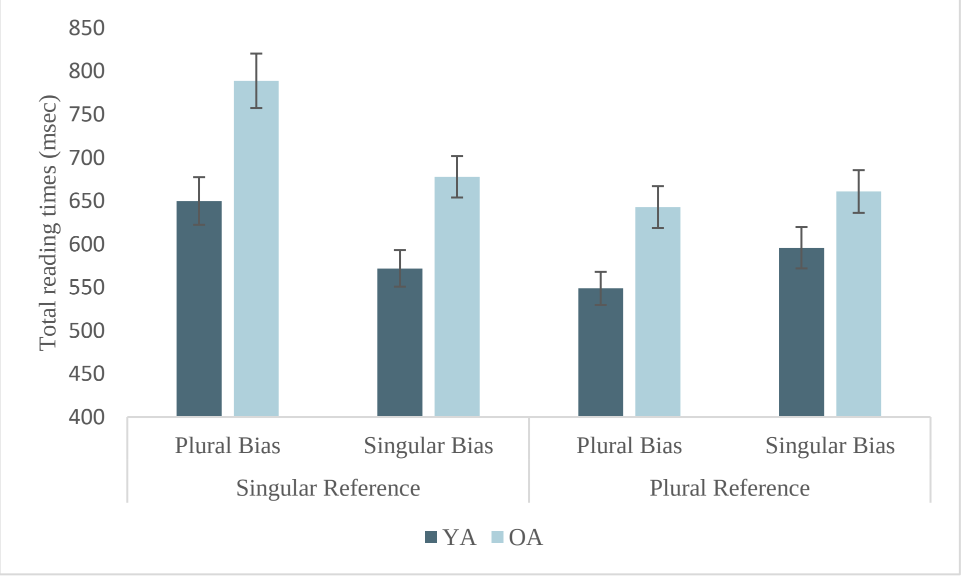
7.3.5. Age effects

Older adults had longer first-pass and total reading times than young adults. Again, this finding is in line with previous research indicating generally longer reading times for older adults. A Reference x Age interaction in first-pass reading times showed that older adults (466 msec) had longer first-pass reading times for singular NP anaphors than young adults (408 msec), whereas there was no age difference for plural NP anaphors (young adults

= 419 msec; older adults = 437 msec). There was also a Context x Reference x Age interaction in total reading times (see Figure 28). The interaction showed that compared to young adults, older adults had longer reading times for all conditions.

Figure 28

Total reading time at the post-critical region



7.4. General discussion

The present study investigated the role of a number of factors in the on-line processing of anaphoric reference to quantified entities during normal reading, with a particular focus on world knowledge, age, and working memory. The results demonstrated a clear and consistent influence of world knowledge on processing across a number of reading time measures in both the critical and post-critical regions of text, with increased processing difficulty when the number of entities referred to by the NP anaphor did not match the bias set up by the context. There was also an influence of age, in that older adults experienced more processing difficulty than young adults in cases where grammatical and pragmatic factors were in conflict with each other. However, working memory showed no observable effects.

These results add to the literature regarding the factors that can influence on-line processing of anaphoric reference to quantified phrases during normal reading. Previous eye-tracking studies which have examined structural and semantic factors in quantifier scope processing, such as the linear order in which the quantifiers appear in the sentence, and intrinsic properties of the quantifiers themselves (e.g., Filik et al., 2004; Paterson et al., 2008) have shown that although these factors exert an influence on reading times on the quantified phrases, participants nevertheless had shorter reading times for subsequent singular than plural anaphoric reference across conditions. The current findings would suggest that in contrast to structural and semantic factors, pragmatic factors (specifically, world knowledge) exert a clear influence on reading times. Intuitively, one can see why contextual bias might play an important role in this case, for instance, if a boutique hotel advertised that "Every room has a bathroom", it would be disappointing to arrive and discover a shared facility. However, it is important to note that there was some evidence of all factors influencing processing in the current study, with more disruption when structural/semantic and pragmatic biases were in opposition to each other. This effect was stronger in the older adult participants compared to the younger participants.

In terms of the previous literature in relation to aging and quantifier scope processing, these results are somewhat unexpected. Based on their results, Kemtes and Kemper (1999) proposed that adults, particularly older adults with a lower working memory span, have a preference for singular NP anaphors following the "principle of singular reference". The current results did not indicate a greater preference for singular continuations in older than younger adults, and there was no evidence of modulation by working memory. It should be noted that differences in methodology between the two studies could mean that Kemtes and Kemper's results are more reflective of post-interpretive processes, rather than processes which occur on-line during natural reading. In contrast, the current findings showed more evidence in favour of young adults having a preference for singular continuations. This might suggest that, young adults adopt a superficial matching strategy between the morphological features of the anaphor and its antecedent (which is always quantified by *a* and therefore always superficially singular), as found in previous studies (Filik et al., 2004; Paterson et al., 2008). For older adults, the present results were more consistent with them initially adopting an underspecified representation in which a commitment to one interpretation of the ambiguity has not been made (Dwivedi, 2013; Raffray & Pickering, 2010). When readers rely on context and world knowledge instead of conducting a full analysis of text, this can result in a partial or underspecified representation of this nature (Sanford & Sturt, 2002). Thus, young and older adults may adopt different strategies when processing this type of ambiguity online, possibly reflecting age groups differences in the use of contextual and world knowledge.

In relation to theories of quantifier scope resolution, in the present experiment, experimental sentences always began with *every* followed by *a*. Thus, following both the principle of linear order (Lakoff, 1971), and the quantifier hierarchy (Ioup, 1975), participants should have showed shorter reading times for plural than singular NP anaphors. Specifically, in terms of linear order, *every* came before *a* in all of the experimental sentences, and in terms of the quantifier hierarchy, *every* is a ‘stronger’ quantifier than *a* – both of which should therefore lead to a preference for a plural referent. However, the present findings indicated that plural NP anaphors resulted in longer reading times.

Our results showed a strong influence of pragmatic bias, evident in a number of measures of reading behaviour across both analysis regions. Despite structural and semantic factors not being manipulated experimentally in the current study, there seemed to be evidence of greater disruption to processing in conditions where structural and semantic cues were in opposition to pragmatic cues, rather than supportive of them. There was also evidence of this disruption being greater for older than younger adults. This finding is in line with accounts that allow for multiple factors or constraints to be considered in parallel (e.g., Kurtzman & MacDonald, 1993). Indeed, this would fit well with a recent drive to develop detailed constraint-satisfaction accounts to explain aspects of higher-level language processing (e.g., Degen & Tanenhaus, 2019). The first step in developing such accounts is to identify and quantify the relevant constraints – the current study would suggest that pragmatic factors are a key constraint in resolving quantifier scope ambiguities, but that structural and semantic factors, as well as individual differences, such as age, may play an additional, if weaker, role.

Although working memory was assessed and entered into the analysis, many of the models did not converge or were not considered a better model than the base model (which did not include working memory). This may reflect Kemtes and Kemper's (1999) findings, which showed that working memory influenced comprehension but not reading times. Thus, working memory may not be a significant factor in online reading but instead may influence later comprehension processes (Caplan et al, 2011; Dede, 2014; Kemtes & Kemper, 1999). Another possibility may be that the sentences did not generate a large enough working memory load to find effects. As we were not fully able to assess the role of working memory in quantifier scope processing, future work is needed to further examine individual difference factors such as working memory, and how they may influence on-line processing of quantifier scope ambiguities.

7.4.1. Conclusion

In conclusion, this study examined the role of a number of factors in processing anaphoric reference to quantified antecedents, with a focus on world knowledge, age, and working memory. Results suggest that pragmatic factors have a clear and consistent influence on processing, with more disruption to reading when the number of entities referred to by an NP anaphor does not match the bias introduced in the context. Pragmatic factors do not influence processing alone, however, with evidence of greater processing difficulty when pragmatic factors are in opposition to structural and semantic factors – particularly for older adults. Overall, results are consistent with a parallel constraint-satisfaction approach.

Chapter 8: General Discussion

8.1. Summary of findings

In this thesis, I aimed to examine whether young and older readers engaged in shallow processing online by conducting four eye-tracking experiments. In these experiments, I also assessed whether age differences in shallow processing between younger and older adults occur at the word-, sentence-, and wider-discourse level. Additionally, the relationship between eye movements during reading and offline comprehension was explored. Finally, working memory and vocabulary assessments were conducted to understand how these factors may modulate shallow processing in young and older adults.

As described in Chapter 1, older adults are of particular interest because across the lifespan, components of fluid intelligence (e.g., memory) deteriorate but elements of crystallised intelligence (e.g., vocabulary) are preserved in later adulthood and our knowledge of the world expands (Paterson et al., 2020; Salthouse, 2009). Evidence from eye-tracking studies suggest older adults may adopt a risky reading strategy, which is where older adults aim to predict what upcoming words will be to ameliorate any natural visual or cognitive decline (McGowan et al., 2015; McGowan et al., 2014; Paterson et al., 2013a; Rayner et al., 2006, 2011, 2013; Zhang et al., 2022). Together, these areas of research suggest that older adults may be more likely to utilise strategies based on their world knowledge during language processing. Chapter 2 described relevant evidence of shallow and good enough processing at the word-, sentence- and wider-discourse level. Young adults' engagement in good enough processing has been widely established. Evidence from Chapter 2 identified that older adults' susceptibility to good enough processing seems to be mixed.

A substantial body of research has focused on measuring language comprehension using offline measures, such as, accuracy of comprehension questions presented immediately after experimental items. Where individuals fail to correctly answer comprehension questions, this has been taken as evidence of a reliance on heuristics rather than algorithmic processing, resulting in a good enough representation. Online measures have typically comprised of self-paced reading tasks or ERPs. In Chapter 3, the advantages of using eye-tracking technology were described. Eye-tracking technology measures the location, duration and sequence of the eye movements while participants read text naturally on a computer screen and so has many benefits compared to self-paced reading tasks and ERPs. This method was deemed the most appropriate methodological technique as it also enables measuring of common eye movement behaviours such as skipping, regressing back to previous parts of the text as well as reading times. Eye-tracking technology was used to investigate the potential age differences in processing of sentences which aimed to assess good enough processing at the word, sentence and wider discourse level (Chapters 4-7). Further analyses of the eye tracking data, along with accuracy rates to comprehension questions enabled the exploration of the relationship between online processing and good enough processing to better understand the underlying processes of language comprehension (Ferreira et al., 2019).

In Chapter 4, Experiment 1 focused on the examination of good enough processing at the word-level by presenting items containing hard-to-detect semantically anomalous words or phrases. At the beginning of the eye-tracking experiment, instructions to young and older participants comprised of reading for comprehension and alerting the experimenter to anything that was ‘out of place’ in the text. After the eye-tracking experiment, a questionnaire was presented to participants which focused on knowledge of the anomalous words presented throughout the experiment. The accuracy of answers to the questionnaire, along with scores of whether an anomaly was identified during the experiment were combined and used to assess whether anomalies had been detected (correctly detected during the experiment and correctly answered in the questionnaire) or undetected (not detected during the experiment but correctly answered in the questionnaire). Analyses consisted of comparisons between detected, undetected and non-anomalous items. The findings from this experiment indicated that young readers were significantly less likely to identify anomalies than older adult readers. When semantic anomalies were successfully detected, young and older adults had longer reading times and made more regressions back to earlier parts of the text. Some evidence of unconscious processing of semantic anomalies was identified during online processing of text, particularly in older adult readers as they made more regressions to specific regions and produced longer regression path and total reading times for undetected anomalies compared to non-anomalous controls. As anomaly detection was taken as an index of shallow processing, the results from this study suggest young adults engaged in shallow processing more than older readers. Both young and older adults resolve anomalies in a similar fashion. The older adult group may have been able to better utilise their linguistic experience to identify hard-to-detect semantic anomalies. A relationship between eye movement behaviour during reading and detection accuracy was found. When participants reprocessed critical regions of text, this predicted greater detection accuracy, thus furthering our knowledge of the relationship between moment-to-moment online processing and comprehension accuracy.

In Chapters 5 and 6, Experiment 2 and 3 investigated age differences in good enough processing at the sentence-level. In Experiment 2, sentences containing temporary syntactic ambiguities were presented to young and older adults (i.e., *While the gentleman was eating the burgers were still being reheated in the microwave*). Additionally, the influence of working memory was investigated by manipulating the amount of text between the ambiguity and subsequent disambiguating information, and by assessing working memory capabilities for both age groups. After reading experimental items, participants were then asked comprehension questions to assess depth of processing (*Was the gentleman eating a burger?*). Reading times were longer at text regions providing syntactic disambiguation ("were still"), semantic disambiguation ("being reheated") and sentence wrap-up ("in the microwave") for ambiguous compared to unambiguous sentences (which contained a comma after "eating") in both young and older adult participants. Young and older adults used different strategies for ambiguity resolution but ultimately achieved similar levels of comprehension. Upon encountering the syntactic (*were still*) and semantic (*being reheated*) disambiguation phrases, young adults immediately attempted reanalysis of the misinterpretation. In contrast, older adults were more likely to wait until the end of the sentence to resolve temporary syntactic ambiguities. Previous research has suggested adult age differences in the comprehension of temporary structural ambiguities during reading. Interestingly, the young and older adults in the current experiment exhibited similar accuracy for the comprehension questions. Accordingly, the present results revealed that processing strategies may differ across adult age groups for ambiguity resolution, however, the ability to resolve structural ambiguities remains intact into later adulthood. This study also furthered our understanding of the relationship between eye movement behaviour and comprehension accuracy, showing that more regressive eye movements are associated with better comprehension accuracy.

When reading single sentences, older adults have been found to skip words more often (McGowan, 2014; McGowan et al., 2015; Paterson et al., 2013a; Rayner et al., 2006; 2011), regress back to earlier parts of text (McGowan et al., 2014, 2015; Paterson et al., 2013b; Rayner et al., 2006; 2011) and make longer forward eye movements (McGowan, 2014; McGowan et al., 2015; Paterson et al., 2013; Rayner et al., 2006; 2011) compared to their younger counterparts. Rayner et al. (2006) suggested that older adults employ a risky reading strategy whereby they aim to predict what upcoming words will be but then when these interpretations are incorrect, they are then more likely to regress back to earlier parts of the text. The consequences of risky reading behaviour for successful text comprehension are currently unknown. Experiment 3 (Chapter 6) aimed to further investigate the relationship between the indices of online processing and offline comprehension in young and older adults to assess whether there is a relationship between risky reading and good enough processing. In this eye-tracking experiment, similar to Ferreira's (2003) study, plausible and implausible events were presented in an active (*The dog bit the man*) and passive sentence structure (*The dog was bitten by the man*) to young and older adults. These examples are another classic example of good enough processing as readers have been found to misinterpret the passive version to mean *the dog bit the man* even though the sentence presents the opposite information (*The man bit the dog*). After each item, a question was presented to participants asking them who the agent (or ‘doer’) was (i.e., *Who did the biting?*). The accuracy scores were used as an index of good enough processing.

The average number of skips across the target sentence, the average progressive saccade length and the number of regressive saccades were measured as characteristics of risky reading. It was predicted that risky reading and shallow processing may be related. The results showed that both young and older adults were found to engage in shallow processing as more errors were made when reading implausible sentences and passive sentences broadly supporting the good enough processing framework. In contrast to predictions, young adults made significantly more errors than older adults. Furthermore, the results from this study furthered support for the risky reading strategy as older adults were found to have a higher rate of skipping, more regressive saccades and on average a longer progressive saccade length for target sentences. Importantly, some evidence for a relationship between risky reading and good enough processing was found as more regressive saccades and more skips predicted more errors on questions. Age differences in the relationship between eye movement behaviour and offline comprehension were identified. Lower rates of accuracy occurred when young adults were likely to skip words and had shorter progressive saccades. For older adults, more regressive saccades predicted more errors on offline comprehension. Overall, these results showed that whilst young and older adults may use different strategies during online processing, offline comprehension remains intact into later adulthood.

In Chapter 7, Experiment 4 explored shallow processing at the discourse level by presenting doubly quantified sentences to young and older adults which biased the reader to either a singular (e.g., *Every student in the class is listed on a register*) or plural (e.g., *Every school pupil in the country has their attendance marked on a register*) entity. A singular (*This register…*) or plural (*These registers…*) continuation then followed. Results showed that contextual bias influenced on-line linguistic processing, with longer reading times for continuations which were incongruent with the number of entities implied by the context. There was also evidence of greater processing difficulty when contextual factors were in opposition with the structural-based preference for the first quantifier in the sentence to take wide scope, especially for older adult readers. Overall, the findings from this study supported a parallel constraint-satisfaction approach under which multiple factors interact during on-line processing.

8.2. Good enough processing

The experiments presented in this thesis aimed to explore age differences in good enough processing at the word-, sentence- and wider discourse level. These studies have found many novel results that have extended the understanding of age differences in good enough processing. In this section, the following will be discussed: overall themes, and implications.

8.2.1. Good enough processing: overall themes and implications

Experiments 1-4 comprised of presenting sentences which have traditionally elicited a good enough interpretation. The prediction was that when reading these types of sentences older adults may be more likely to utilise context and their world knowledge, thus engaging in good enough processing more than young adults when reading sentences (Christianson et al., 2006; Dede, 2014; Kemper et al., 2004; Kemtes & Kemper, 1997, 1999; Malyutina & den Ouden, 2016). The rationale for older adults being more likely to engage in good enough processing came from literature showing that this age group encounter greater reading difficulty (Kemper & McDowd, 2006; Kliegl et al., 2004; McGowan et al., 2015; McGowan et al., 2014; Paterson et al., 2013a, 2013b; Rayner et al., 2006, 2011, 2013; Whitford & Titone, 2017, 2016) and that there is greater deterioration in elements of fluid intelligence during aging (Paterson et al., 2020; Salthouse, 2009). In contrast, across the lifespan elements of crystallised intelligence stabilise and can expand (Ben-David et al., 2015; Brysbaert et al., 2016; Keuleers et al., 2015). Therefore, it was expected that older adults may engage in good enough processing, utilising their linguistic and world knowledge more than young adults to compensate for deterioration in elements of fluid intelligence. The findings from this thesis highlighted that comprehension is preserved in later adulthood. Interestingly, despite older adults' greater linguistic and world knowledge, this age group did not engage in good enough processing more so than their younger counterparts. This finding has implications for the good enough account which stipulates that readers will employ fast and frugal heuristic to derive an interpretation which will be good enough for the given task (Ferreira et al., 2002; Karimi & Ferreira, 2016). Importantly, the results suggest that those who have greater linguistic and world knowledge may not be any more likely to engage in good enough processing.

Another important implication is that the good enough processing framework cannot alone explain all of the results from the experiments presented in this thesis. For instance, at the word-level (Experiment 1), evidence of unconscious processing of undetected semantic anomalies was identified. The occurrence of unconscious processing is in contrast with the predictions of the shallow processing accounts (Ferreira et al., 2002; Sanford & Sturt, 2002). Sanford et al., (2011) suggested that semantic anomalies occur due to a fault in the retrieval or integration of a word meaning with the wider discourse, this results in the anomaly not being registered in the reader's comprehension system. However, the results from Experiment 1 showed evidence of unconscious processing as differences between undetected anomalies and non-anomalous sentences were found, particularly in older adults who were more likely to have longer regression path and total reading times for undetected sentences compared to non-anomalous sentences. Instead, the findings from Experiment 1 were consistent with the RI-Val model (Cook et al., 2016). This model explains that in order for semantic anomalies to be detected, incoming text must be integrated with a readers world knowledge (resonance phase), this information can then be integrated with contextual information from the text, which is contained in the working memory (integration phase), this representation is then validated (validation phase) (Cook et al., 2016). A representation is achieved when the interpretation meets the coherence threshold. Cook et al. (2016) suggested semantic anomalies that go undetected result in delayed processing and that the information that would discriminate the anomaly is not successfully completed in either the resonance, integration or validation stage.

At the sentence level, evidence of good enough processing was found in young and older readers when processing sentences containing temporary syntactic ambiguities but older adults were not significantly more likely to engage in good enough processing than their younger counterparts (Experiment 2). Predictions would propose that older adults would be more likely to engage in good enough processing, possibly due to a failure in erasing the incorrect initial misinterpretation or in integrating the new syntactic and semantic information. In this study, the length between the ambiguous information and the disambiguating information was manipulated to understand if the longer the reader holds to an incorrect interpretation results in the likelihood that the reader engages in good enough processing. The results from this study showed that older adults were likely to reprocess shorter ambiguous sentences but encountered more difficulty in longer ambiguous sentences, which may support evidence suggesting that working memory may influence engagement in good enough processing.

In Experiment 3, young and older adults made more errors when reading passive sentences (compared to active structures) and for implausible sentences (compared to plausible sentences). However, the results did not entirely replicate Ferreira's (2003) findings as no interaction between plausibility and structure was found. Whilst these findings can be attributed to good enough processing, they can also be explained by the retrieval account put forward by Bader and Meng (2018; Meng & Bader, 2021). The retrieval account advocates that errors are due to difficulty in retrieving information which is required to correctly answer questions pertaining to thematic roles, thus it is simply due to task demands (Bader and Meng, 2018; Meng & Bader, 2021).

At the discourse level, the results indicated that context was considered during online processing of doubly quantified sentences, however, there were multiple factors interacting, instead supporting a parallel constraint-satisfaction approach. Furthermore, age differences occurred with older adults underspecifying interpretations more so than young adults.

8.3. Eye movement data and aging

Four eye-tracking experiments were conducted to examine the age differences in online processing of sentences that traditionally result in a good enough interpretations. These experiments have made a clear contribution to understanding age differences in the online processing of sentences often eliciting a good enough interpretation. In this section, overall themes, implications and future directions are discussed.

8.3.1. Eye movement data and aging: overall themes, implications and future directions

Few studies have used eye-tracking to examine the online processing of sentences which traditionally result in good enough representations. The results from the experiments in this thesis suggest that young and older adults employ different strategies during sentence processing with older adults requiring additional time to process complex sentences but correct interpretations follow in offline comprehension. In contrast, young adults are more likely to conduct a reanalysis immediately.

8.3.2. Risky reading

Since Rayner et al.'s (2006) influential paper, which re-established the risky reading strategy (O'Regan, 1990), this strategy has been a leading hypothesis for the differences identified during reading between young and older adult readers. Choi et al. (2017) argued that older adults do not engage in a risky reading strategy. However, a recent meta-analysis has shown that older adults of alphabetic languages do engage in a risky reading strategy (Zhang et al., 2022). Experiment 3 provides a clear contribution to the risky reading literature by extending knowledge of the instances in which older adults engage in risky reading. In Experiment 3, multi-line sentences were presented where plausibility and sentence structure were manipulated. Then a question followed after every item. Identifying a risky reading strategy under these conditions contrasts with previous evidence which has suggested that older adults produce a more cautious strategy when a question is presented after every item (Woltschak & Kliegl, 2013).

The underlying mechanisms of risky reading are currently unknown. The meta-analysis conducted points to non-linguistic reasons (Zhang et al., 2022), such as, poor saccadic control or that older adults engage in mindless reading or focus on understanding the general discourse and focus less on the details. Whatever the underlying reasons are, the characteristics of the risky reading strategy are that older adults tend to skip words more often (McGowan, 2014; McGowan et al., 2015; Paterson et al., 2013a; Rayner et al., 2006; 2011), regress back to earlier parts of text (McGowan et al., 2014, 2015; Paterson et al., 2013b; Rayner et al., 2006; 2011) and make longer forward eye movements (McGowan, 2014; McGowan et al., 2015; Paterson et al., 2013; Rayner et al., 2006; 2011). Experiment 3 focused on the relationship between offline comprehension and risky reading and found a relationship. Future work could consider other linguistic consequences of risky reading.

8.4. The relationship between online processing and offline comprehension

Further analyses comprised of conducting a logistic regression to analyse the relationship between online processing and offline comprehension. Eye movements were measured as an index of online processing and accuracy was an index of offline comprehension. The results from these analyses have also made a clear contribution to the understanding of the relationship between online processing and offline comprehension. This section focuses on overall themes, implications and future directions.

8.4.1. The relationship between online processing and offline comprehension: overall themes, implications and future directions

Few studies have investigated the relationship between online processing and offline comprehension (Christianson & Luke, 2011; Huang & Ferreira, 2021; Qian et al., 2018; Wonnacott et al., 2016) and even fewer have done so using eye-tracking technology (Huang & Ferreira, 2021; Wonnacott et al., 2016) and with different age groups (Howman & Filik, 2020; Wonnacott et al., 2016). The results from this thesis found that the following eye movements predicted offline comprehension; more regressions, longer total reading times, shorter first-pass reading times and higher probability of skipping.

The association between eye movements and offline comprehension is becoming a topical issue to explore (Christianson & Luke, 2011; Ferreira et al., 2019; Howman & Filik, 2020; Huang & Ferreira, 2021; Qian et al., 2018; Wonnacott et al., 2016). Additional studies measuring the eye movements that are linked to offline comprehension would be welcome, especially exploring how this may differ when accounting for individual differences to further understand the relationship between eye movements during reading and offline comprehension.

Furthermore, some of the characteristics of risky reading; more skipping and more regressive saccades predicted more errors. Despite the underlying mechanism of risky reading currently unknown, the findings in this thesis have advanced the understanding of the behaviour of those that engage in risky reading as the results from Experiment 3 suggest that some of the behaviours of risky reading may be associated with poorer comprehension. It is beyond the scope of this thesis, however, these findings have potential implications for models of reading comprehension, which may in future want to consider the relationship between eye movements during online processing of text and ultimate comprehension.

The studies conducted in this thesis indicate the usefulness of using eye-tracking methodology when investigating good enough processing. Eye-tracking technology records the moment-to-moment eye movements. The utilisation of eye-tracking methodology enabled the identification of a key finding in that young and older adults use different strategies used during online processing.

Furthermore, analysing the relationship between eye movements during reading and offline comprehension has advanced understanding in the mechanisms underlying offline comprehension. Previously, studies have found no relationship between key eye movement measures, such as regressions and a relationship with total reading times has been mixed (Christianson & Luke, 2011; Howman & Filik, 2020; Huang & Ferreira, 2021; Qian et al., 2018; Wonnacott et al., 2016). The studies in this thesis show that good-enough processing occurs online and that there is a relationship with this and offline comprehension.

The few studies that have explored the relationship between eye movements during reading and offline comprehension have predominantly used comprehension questions. However, whether eye movements are related to other measures of offline comprehension (i.e., cloze tests), inference, gist, recall, paraphrasing tasks would be an interesting avenue to explore for a complete understanding of the relationship between eye movements during reading and offline comprehension. This is especially pertinent as recently the types of comprehension questions have been criticised for only focusing on the initial misanalysis (Chromý, 2022).

Four eye-tracking experiments were conducted for this thesis. Eye-tracking is an appropriate method to investigate risky reading and age differences in the online processing of good enough processing as this technology records the moment-to-moment eye movements. However, a limitation of eye-tracking is that neural activity cannot be measured. Neural activity is particularly important as ERP studies have found both the P400 and N600 to be associated with good enough processing (see Chapter 3). The issue with ERP studies is that items are presented similarly to self-paced read tasks, a word or phrase at a time. The co-registration of eye movements and ERPs is a new technique that has emerged (see Degno et al., 2021 for more information on the methods) and can address this issue. Novel results have been found for young adults but currently there are no published studies with older adults. This would be an interesting technique to explore in the future to further understand the underlying processes of good enough processing in young and older adults.

8.5. Limitations and future directions

One explanation for the difference in accuracy scores in the experiments described in this thesis compared to previous findings (Christianson et al., 2006; Daneman et al., 2006; Dede, 2014; Kemtes & Kemper, 1997, 1999; Malyutina & den Ouden, 2016) may be that task demands contributed to the findings. Comprehension questions were presented after every item in Experiments 1 and 3. For Experiments 2 and 4, comprehension questions were presented for 50% of items in the overall experiment, however for the experimental items in Experiment 2, there was also a question after every item. Older adults may have adapted to the task demands of the experiment as older adults were less likely to engage in 'good enough' processing in Experiments 1 and 3. There were no age differences in Experiments 2 and 4. Previously, it has been suggested that readers may engage in more careful reading when there is a question after every item (Wotschack & Kliegl, 2013), this could have led older adults to place a focus on comprehension in Experiment 1 and 3 more than they would typically in other instances.

Previously, explicit instructions on anomaly detection tasks has resulted in different accuracy rates (Bohan, 2007). In contrast, when participants were asked about the explicit wording of garden-path sentences, for instance, '*Did the sentence explicitly say…*', Qian et al.

(2018) found no differences on comprehension. Given these mixed findings on task demands, coupled with the differences in older adults' likelihood to engage in good enough processing, exploring age differences in the effects of task demands, such as, how instruction type and question type influence depth of processing may be an interesting next step. Understanding task demands could have important applications as there are numerous instances when individuals are required to read for high accuracy such as exam questions, news articles and legal documents. Previously, drug name confusion has been reduced by using particular fonts (Filik et al., 2006). This example illustrates how amending salient information can reduce errors on word recognition. Similarly, future studies could examine task demands to understand what may influence age differences in good enough processing.

A limitation of the experiments is that older adult participants were very enthusiastic about their participation, in contrast most of the young adult participants were psychology undergraduates who participated in exchange for course credit or a small monetary honorarium. No measure of motivation was measured during the experiments, however, this may have also influenced accuracy scores.

Christianson et al. (2006) suggested older adults are more likely to derive a blended interpretation than young adults. This was beyond the scope of this thesis but is a pertinent consideration. Whilst Malyutina and den Ouden (2016) aimed to explore this with a picture matching task, due to ceiling effects this was not successfully investigated. The findings in this thesis seem to suggest that older adults have difficulty integrating the information together. This hypothesis could be tested in the future.

Despite recruiting large samples and using R Studio, a powerful statistical programme, a limitation is that many of the models including working memory and vocabulary covariates did not converge. Furthermore, young and older adults had similar working memory and vocabulary assessment scores. Most likely, for these reasons, the author was unable to investigate the role of working memory and vocabulary on good enough processing in young and older adults. One consideration is that literature has widely documented that aging is associated with cognitive decline. However, “Super Agers” are older adults who exhibit preserved memory function (Corrada, 2021; Harrison et al., 2012; Nicholls, 2022; Richmond et al., 2022). This may explain why older adult participants in all four experiments on average had better working memory scores when compared to their younger counterparts, despite literature suggesting otherwise. The little variation in scores between groups may also explain why the models including working memory and vocabulary did not result in differences.

The WAIS digit forward, backward, and sequencing tasks is one of the most widely used measures of working memory. However, the separate language interpretation resource theory (e.g., Caplan & Waters, 1990, 1999a, 1999b; Waters & Caplan, 1996) advocates that a domain-specific vWM system exists which is integral to online language comprehension. According to this theory, common WM measures, such as digit span and the Daneman and Carpenter (1980) reading span task may not be associated with all measures of language processing. More data may have been required to interrogate the influence of working memory and vocabulary on good enough processing. Furthermore, using a battery measures which assess different factors of fluid intelligence would be beneficial as processing speed and inhibitory control have been linked to working memory (Salthouse & Meinz, 1995) and good enough processing (Anema & Obler, 2012; Christianson et al., 2006; Dede, 2014; DeDe et al., 2004; Kemper et al., 2004; Kemtes & Kemper, 1997; Slattery et al., 2013). This would be an interesting next step to understand the sensitivities between different cognitive factors and good enough processing and could have important implications for the good enough processing framework as currently working memory is considered a critical reason for individual differences (Karimi & Ferreira, 2016).

Another noteworthy point is that trimming of the data was conducted for young and older adults. Removing outliers in experiments is a long-standing issue in scientific research (Pollet & Van Der Meij, 2017). Removal of outliers can impact parameter estimates when using maximum likelihood estimates (Aguinis et al., 2013) leading to a risk of Type 1 or Type II error. Trimming outliers ± 2.5 standard deviations from the mean is typically performed in eye-tracking experiments (Cop et al., 2017; Fitzsimmons et al., 2020; Hermena et al., 2017) and this was the approach taken in the experiments in the thesis. However, recent evidence suggests that this may be problematic (Andre, 2021). Data loss was similar per condition per group and is a common method to removal of outliers (Pollet & Van Der Meij, 2017).

A limitation of the experiments in thesis is that there may have been low power. Conducting power analyses for LMEMs is difficult as power formulas do not typically account for a sample of items to a sample of participants, Brysbaert & Stevens' (2016) influential article advocated that 1600 observations would be needed for a well powered study. This was the aim for the experiments, and the sample sizes were as big and in many instances much larger than previous experiments in this area. More recently, more simulations have been devised and would be used for future studies (e.g., Judd et al., 2016).

A final limitation is that a correction for multiple comparisons were not conducted. Recently, Bonferroni correction has been recommended (Von der Malsburg et al., 2017), however as eye movements are correlated, this may not be the most appropriate correction for eye-tracking experiments. The experiments in this thesis had a clear experimental design, which were theoretically driven and as no priori was set and there is no agreement in the field, it was not deemed appropriate to decide on corrections after seeing the results. In future, pre-registering studies and agreeing on the approach a priori would be the best approach.

8.6. Conclusion

In this thesis, I examined whether young and older adults engage in shallow processing online. Four eye-tracking experiments were conducted to examine whether age differences in shallow processing between younger and older adults occur at the word-, sentence-, and wider-discourse level. Additionally, the relationship between eye movements during reading and offline comprehension was explored.

Shallow processing and good enough processing are two theoretical accounts which advocate that a full syntactic and semantic analysis may not always be accomplished, instead readers often rely on heuristics, such as context and word knowledge, which leads to an underspecified, shallow or good enough representation (Ferreira et al., 2002; Sanford & Graesser, 2006; Sanford & Sturt, 2002). Good enough processing has been widely investigated in a young adult population. Older readers were an interesting population to examine good enough processing on as this group may be more inclined to rely on heuristics, such as, linguistic and world knowledge, compared to their younger counterparts to compensate for deterioration of elements of fluid intelligence. This thesis comprised of four eye-tracking experiments which examined the online processing of young and older adults whilst reading sentences that could bias them to a good enough interpretation at the word-, sentence- and wider discourse-level. In Experiment 1, sentences containing hard-to-detect anomalies were presented to assess shallow processing at the word-level. In Experiment 2, sentences containing temporary syntactic ambiguities were presented and Experiment 3 presented plausible and implausible sentences in active or passive sentence structure to assess good enough processing at the sentence-level. Finally, to examine the discourse-level, doubly quantified sentences were presented to young and older adults. There were several novel findings. In contrast to the prediction of the author, older adults were no more likely to engage in good enough processing than younger readers. Eye-tracking results revealed that young and older adults may use different strategies during online processing of ambiguities but ultimately achieve similar levels of comprehension (Experiments 2 and 3). The relationship between eye movements and accuracy was examined in Experiment 1, 2 and 3 and various eye movements were found to predict good enough processing, especially regressions. Finally, whilst shallow and good enough processing accounts can account for some of the findings, the results suggest that these frameworks cannot explain all of the findings. The findings from the experiments conducted in thesis have raised numerous potential future directions which could be explored further to better understand age differences in good enough processing.

References

Aaronson, D. (1976). Performance theories for sentence coding: Some qualitative observations. *Journal of Experimental Psychology: Human Perception and Performance*, *2*(1), 42–55. https://doi.org/10.1037/0096-1523.2.1.42

Aguinis, H., Gottfredson, R. K., & Joo, H. (2013). Best-practice recommendations for defining, identifying, and handling outliers. *Organizational Research Methods*, *16*(2), 270-301.

Altmann, L. J. P., & Kemper, S. (2006). Effects of age, animacy and activation order on sentence production. In *Language and Cognitive Processes* (Vol. 21, Issues 1–3). https://doi.org/10.1080/01690960544000006

Anderson, R. S. (2017). Improving ophthalmic diagnosis in the clinic using the Moorfields Acuity Chart. *Expert Review of Ophthalmology*, *12*(6), 433–435. https://doi.org/10.1080/17469899.2017.1395696

Anema, I., & Obler, L. K. (2012). Hyphens for disambiguating phrases: Effectiveness for young and older adults. *Reading and Writing*, *25*(9), 2091–2106. https://doi.org/10.1007/s11145-011-9345-4

André, Q. (2021). Outlier exclusion procedures must be blind to the researcher's hypothesis. *Journal of Experimental Psychology: General*.

Artuso, C., & Belacchi, C. (2021). Semantic memory and reading comprehension: the relationship through adulthood and aging. *Aging Clinical and Experimental Research*, 1–11. https://doi.org/10.1007/s40520-020-01771-1

Au-Yeung, S. K., Kaakinen, J. K., Liversedge, S. P., & Benson, V. (2018). Would adults with autism be less likely to bury the survivors?: An eye movement study of anomalous text reading. *Quarterly Journal of Experimental Psychology*, *71*(1 Special Issue), 280–290.

https://doi.org/10.1080/17470218.2017.1322621

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

Baddeley, A. D., Hitch, G. J., & Allen, R. J. (2009). Working memory and binding in sentence recall. *Journal of Memory and Language*, *61*(3), 438–456. https://doi.org/10.1016/j.jml.2009.05.004

Bader, M., & Meng, M. (2018). The misinterpretation of noncanonical sentences revisited. *Journal of Experimental Psychology: Learning Memory and Cognition*, *44*(8), 1286–1311. https://doi.org/10.1037/xlm0000519

Balota, D. A., Pollatsek, A., & Rayner, K. (1985). The interaction of contextual constraints and parafoveal visual information in reading. *Cognitive Psychology*, *17*(3), 364–390. https://doi.org/10.1016/0010-0285(85)90013-1

Balota, D. A., Yap, M. J., & Cortese, M. J. (2006). Visual word recognition : the journey from features to meaning. In *Word Journal Of The International Linguistic Association*.

Baltes, P. B., Staudinger, U. M., & Lindenberger, U. (1999). Lifespan psychology: Theory and application to intellectual functioning. *Annual Review of Psychology*, *50*, 471–507. https://doi.org/10.1109/ICIS.2005.28

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

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

Barton, S. B., & Sanford, A. J. (1993). A case study of anomaly detection: Shallow semantic processing and cohesion establishment. *Memory & Cognition*, *21*(4), 477–487. https://doi.org/10.3758/BF03197179

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Bellocchi, S., Muneaux, M., Bastien-Toniazzo, M., & Ducrot, S. (2013). I can read it in your eyes: What eye movements tell us about visuo-attentional processes in developmental dyslexia. *Research in Developmental Disabilities*, *34*(1), 452–460. https://doi.org/10.1016/j.ridd.2012.09.002

Ben-David, B. M., Erel, H., Goy, H., & Schneider, B. A. (2015). "Older is always better": Age-related differences in vocabulary scores across 16 years. *Psychology and Aging*, *30*(4), 856–862. https://doi.org/10.1037/pag0000051

Bohan, J., & Filik, R. (2018). Perspective effects during reading: Evidence from text change-detection. *Discourse Processes*, *55*(2), 113-122.

Bohan, J., Leuthold, H., Hijikata, Y., & Sanford, A. J. (2012). The processing of good-fit semantic anomalies: An ERP investigation. *Neuropsychologia*, *50*(14), 3174–3184. https://doi.org/10.1016/j.neuropsychologia.2012.09.008

Bohan, J., & Sanford, A. J. (2008). Semantic anomalies at the borderline of consciousness: An eye-tracking investigation. *Quarterly Journal of Experimental Psychology*, *61*(2), 232–239. https://doi.org/10.1080/17470210701617219

Bohan, J. T. (2007). Depth of processing and semantic anomalies. *(Doctoral Dissertation, University of Glasgow)*. http://theses.gla.ac.uk/127/

Borella, E., Carretti, B., & De Beni, R. (2008). Working memory and inhibition across the adult life-span. *Acta Psychologica*, *128*(1), 33–44. https://doi.org/10.1016/j.actpsy.2007.09.008

Borella, E., Ghisletta, P., & De Ribaupierre, A. (2011). Age differences in text processing: The role of working memory, inhibition, and processing speed. *Journals of Gerontology - Series B Psychological Sciences and Social Sciences*, *66 B*(3), 311–320. https://doi.org/10.1093/geronb/gbr002

Brédart, S., & Docquier, M. (1989). The Moses illusion: A follow-up on the focalization effect. *Current Psychology of Cognition*, *9*.

Bredart, S., & Modolo, K. (1988). Moses strikes again: Focalization effect on a semantic illusion. *Acta Psychologica*, *67*(2), 135–144. https://doi.org/10.1016/0001-6918(88)90009-1

Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of cognition*, *1*(1).

Brysbaert, M., Stevens, M., Mandera, P., & Keuleers, E. (2016). How many words do we know? Practical estimates of vocabulary size dependent on word definition, the degree of language input and the participant's age. *Frontiers in Psychology*, *7*(JUL), 1–11. https://doi.org/10.3389/fpsyg.2016.01116

Burke, D. M., MacKay, D. G., & James, L. E. (2000). Theoretical approaches to language and aging.

Caplan, D., DeDe, G., Waters, G., Michaud, J., & Tripodis, Y. (2011). Effects of age, speed of processing, and working memory on comprehension of sentences with relative clauses. *Psychology and Aging*, *26*(2), 439–450. https://doi.org/10.1037/a0021837

Chen, A. H., Khalid, N. M., & Buari, N. H. (2019). Age factor affects reading acuity and reading speed in attaining text information. *International Journal of Ophthalmology*, *12*(7), 1170–1176. https://doi.org/10.18240/ijo.2019.07.19

Choi, W., Lowder, M. W., Ferreira, F., Swaab, T. Y., & Henderson, J. M. (2017). Effects of word predictability and preview lexicality on eye movements during reading: A comparison between young and older adults. *Psychology and Aging*, *32*(3), 232–242. https://doi.org/10.1037/pag0000160

Chomsky, N. (1965). Persistent topics in linguistic theory. *Diogenes*, *13*(51), 13-20.

Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology*, *42*(4), 368–407. https://doi.org/10.1006/cogp.2001.0752

Christianson, K., & Luke, S. G. (2011). Context strengthens initial misinterpretations of text. *Scientific Studies of Reading*, *15*(2), 136–166. https://doi.org/10.1080/10888431003636787

Christianson, K., Luke, S. G., & Ferreira, F. (2010). Effects of plausibility on structural priming. *Journal of Experimental Psychology: Learning Memory and Cognition*, *36*(2), 538–544. https://doi.org/10.1037/a0018027

Christianson, K., Williams, C. C., Zacks, R. T., & Ferreira, F. (2006). Younger and older adults' "good-enough" interpretations of garden-path sentences. *Discourse Processes*, *42*(2), 205–238. https://doi.org/10.1207/s15326950dp4202_6

Chromý, J. (2022). When readers fail to form a coherent representation of garden-path sentences. *Quarterly Journal of Experimental Psychology*, *75*(1), 169–190. https://doi.org/10.1177/17470218211037152

Clifton, C., Staub, A., & Rayner, K. (2007). Eye movements in reading words and sentences. *Eye Movements*, 341–371. https://doi.org/10.1016/B978-008044980-7/50017-3

Cook, A. E., & O'Brien, E. J. (2014). Knowledge activation, integration, and validation during narrative text comprehension. *Discourse Processes*, *51*(1–2), 26–49. https://doi.org/10.1080/0163853X.2013.855107

Cook, A. E., Walsh, E. K., Bills, M. A. A., Kircher, J. C., & O'Brien, E. J. (2016). Validation of semantic illusions independent of anomaly detection: Evidence from eye movements.

Quarterly Journal of Experimental Psychology, *71*(1 Special Issue), 113–121. https://doi.org/10.1080/17470218.2016.1264432

Cop, U., Dirix, N., Drieghe, D., & Duyck, W. (2017). Presenting GECO: An eyetracking corpus of monolingual and bilingual sentence reading. *Behavior research methods*, *49*(2), 602-615.

Craik, F. I. (1986). A functional account of age differences in memory. *Human memory and cognitive capabilities: Mechanisms and performances*, *5*, 409-422.

Cutter, M. G., Paterson, K. B., & Filik, R. (2022). Online representations of non-canonical sentences are more than good-enough. *Quarterly Journal of Experimental Psychology, 75*(1), 30-42. https://doi.org/10.1177/17470218211032043

Dabrowska, E., & Street, J. (2006). Individual differences in language attainment : Comprehension of passive sentences by native and non-native English speakers. *Language Sciences*, *28*(6), 604–615. https://doi.org/10.1016/j.langsci.2005.11.014

Dambacher, M., Dimigen, O., Braun, M., Wille, K., Jacobs, A. M., & Kliegl, R. (2012). Stimulus onset asynchrony and the timeline of word recognition: Event-related potentials during sentence reading. *Neuropsychologia*, *50*(8), 1852–1870. https://doi.org/10.1016/j.neuropsychologia.2012.04.011

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and aging. *Journal of Verbal Learning and Verbal Behavior*, *19*(4), 450–466. https://doi.org/10.4324/9781315879840-7

Daneman, M., Hannon, B., & Burton, C. (2006a). Are there age-related differences in shallow semantic processing of text? Evidence from eye movements. *Discourse Processes*, *42*(2), 177–203. https://doi.org/10.1207/s15326950dp4202_5

Daneman, M., Hannon, B., & Burton, C. (2006b). Shallow Discourse Processing. *Discourse Processes*, *42*(2), 177–203. https://doi.org/10.1207/s15326950dp4202

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin and Review, 3*(4), 422–433. https://doi.org/10.3758/BF03214546

De Beni, R., Palladino, P., Borella, E., & Lo Presti, S. (2003). Reading comprehension and aging: Does an age-related difference necessarily mean impairment? *Aging Clinical and Experimental Research*, *15*(1), 67–76. https://doi.org/10.1007/BF03324482

Dede, G. (2014). Sentence comprehension in older adults: Evidence for risky processing strategies. *Experimental Aging Research*, *40*(4), 436–454. https://doi.org/10.1080/0361073X.2014.926775

Dede, G., & Flaxx, J. K. (2016). Language comprehension in aging. In *In Wright, H.H. (Eds.), Cognition, language and aging (PP.107-133). Benjamins Publishing Comapny.* https://doi.org/10.1080/03610738508259280

DeDe, G., Kemtes, K., Caplan, D., & Waters, G. (2004). The relationship between age, verbal working memory, and language comprehension. *Psychology and Aging*, *19*(4), 601–616. https://doi.org/10.1037/0882-7974.19.4.601

Degen, J., Tanenhaus, M. K., Degen, J., & Tanenhaus, M. K. (2019). Constraint-based pragmatic processing. *The Oxford Handbook of Experimental Semantics and Pragmatics*, 20–38. https://doi.org/10.1093/oxfordhb/9780198791768.013.8

Degno, F., Loberg, O., & Liversedge, S. P. (2021). Co-Registration of eye movements and fixation—related potentials in natural reading: Practical issues of experimental design and data analysis. *Collabra: Psychology*, *7*(1), 1–27. https://doi.org/10.1525/collabra.18032

Drieghe, D., Veldre, A., Fitzsimmons, G., Ashby, J., & Andrews, S. (2019). The influence of number of syllables on word skipping during reading revisited. *Psychonomic Bulletin and Review*, *26*(2), 616–621. https://doi.org/10.3758/s13423-019-01590-0

Duchowski, A. T. (2003). Eye tracking techniques. In *Eye tracking methodology: Theory and practice* (pp. 55-65). Springer, London.

Duffy, S. A., Morris, R. K., & Rayner, K. (1988). Lexical ambiguity and fixation times in reading. *Journal of Memory and Language*, *27*(4), 429–446. https://doi.org/10.1016/0749-596X(88)90066-6

Dwivedi, V. D. (2013). Interpreting quantifier scope ambiguity: Evidence of heuristic first, algorithmic second processing. *PLoS ONE*, *8*(11), 1–20. https://doi.org/10.1371/journal.pone.0081461

Dwivedi, V. D., & Gibson, R. M. (2017). An ERP investigation of quantifier scope ambiguous sentences: Evidence for number in events. *Journal of Neurolinguistics*. https://doi.org/10.1016/j.jneuroling.2016.11.006

Dwivedi, V. D., Phillips, N. A., Einagel, S., & Baum, S. R. (2010). The neural underpinnings of semantic ambiguity and anaphora. *Brain Research*, *1311*, 93–109. https://doi.org/10.1016/j.brainres.2009.09.102

Elliot, D. B., Yang, K. C. H., & Whitaker, D. (1995). Visual acuity changes throughout adulthood in normal, healthy eyes: Seeing beyond 6/6. In *Optometry and Vision Science* (Vol. 72, Issue 3, pp. 186–191). https://doi.org/10.1097/00006324-199503000-00006

Erickson, T. D., & Mattson, M. E. (1981). From words to meaning: A semantic illusion. *Journal of Verbal Learning and Verbal Behvaiour*, *20*, 540–551.

Faber, M., Mak, M., & Willems, R. M. (2020). Word skipping as an indicator of individual reading style during literary reading. *Journal of Eye Movement Research*, *13*(3). https://doi.org/10.16910/jemr.13.3.2

Federmeier, K. D., & Kutas, M. (2005). Aging in context: Age-related changes in context use during language comprehension. *Psychophysiology*, *42*(2), 133–141. https://doi.org/10.1111/j.1469-8986.2005.00274.x

Federmeier, K. D., & Kutas, M. (2019). Neuropsychologia what ' s " left " ? Hemispheric sensitivity to predictability and congruity during sentence reading by older adults. *Neuropsychologia*, *133*(August), 107173. https://doi.org/10.1016/j.neuropsychologia.2019.107173

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology*, *47*(2), 164–203. https://doi.org/10.1016/S0010-0285(03)00005-7

Ferreira, F., Bailey, K. G. D., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, *11*(1), 11–15. https://doi.org/10.1111/1467-8721.00158

Ferreira, F., & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language*, *25*(3), 348–368. https://doi.org/10.1016/0749-596X(86)90006-9

Ferreira, F., & Henderson, J. M. (1991). Recovery from misanalyses of garden-path sentences. *Journal of Memory and Language*, *30*(6), 725–745. https://doi.org/10.1016/0749-596X(91)90034-H

Ferreira, F., Yang, Z., Ferreira, F., & Yang, Z. (2019). The problem of comprehension in psycholinguistics. *Discourse Processes*, *0*(00), 1–11. https://doi.org/10.1080/0163853X.2019.1591885

Ferris, F. L., & Bailey, I. (1996). Standardizing the measurement of visual acuity for clinical research studies: Guidelines from the Eye Care Technology Forum. *Ophthalmology (Rochester, MN)*, *103*(1), 181-182.

Filik, R., & Leuthold, H. (2008). Processing local pragmatic anomalies in fictional contexts: Evidence from the N400. *Psychophysiology*, *45*(4), 554–558. https://doi.org/10.1111/j.1469-8986.2008.00656.x

Filik, R., Paterson, K. B., & Liversedge, S. P. (2004). Processing doubly quantified sentences: Evidence from eye movements. *Psychonomic Bulletin and Review*, *11*(5), 953–959. https://doi.org/10.3758/BF03196727

Filik, R., Purdy, K., Gale, A., & Gerrett, D. (2006). Labeling of medicines and patient safety: Evaluating methods of reducing drug name confusion. *Human Factors*, *48*(1), 39–47. https://doi.org/10.1518/001872006776412199

Fillenbaum, S. (1971). On coping with ordered and unordered conjunctive sentences. *Journal of Experimental Psychology: General*, *87*(1).

Fillenbaum, S. (1974). Pragmatic normalization: Further results for some conjunctive and disjunctive sentences. *Journal of Experimental Psychology*, *102*(4), 574–578. https://doi.org/10.1037/h0036092

Fitzsimmons, G., Jayes, L. T., Weal, M. J., & Drieghe, D. (2020). The impact of skim reading and navigation when reading hyperlinks on the web. *PloS one*, *15*(9), e0239134.

Folstein, M, F., Folsteing, S, E., & McHugh, P, R. (1975). *"Mini-Mental State Exam". A practical state method for grading the cognitive state of patients for the clinician. 12*, 189–198.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, *14*(2), 178–210. https://doi.org/10.1016/0010-0285(82)90008-1

Frazier, L., & Rayner, K. (1990). Taking on semantic commitments: Processing multiple meanings vs. multiple senses. *Journal of Memory and Language, 29*(2), 181–200. https://doi.org/10.1016/0749-596X(90)90071-7

Friederici, A. D., Pfeifer, E., & Hahne, A. (1993). Event-related brain potentials during natural speech processing: effects of semantic, morphological and syntactic violations. *Cognitive Brain Research*, *1*(3), 183–192. https://doi.org/10.1016/0926-6410(93)90026-2

Gibson, E., Bergen, L., & Piantadosi, S. T. (2013). Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(20), 8051–8056. https://doi.org/10.1073/pnas.1216438110

Goldberg, J. H., & Wichansky, A. M. (2003). Eye Tracking in Usability Evaluation. A Practitioner's Guide. *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research, October*, 493–516. https://doi.org/10.1016/B978-044451020-4/50027-X

Goldstein, D. G., & Gigerenzer, G. (2002). Models of ecological rationality : the recognition heuristic. *Psychological Review*, *109*(1), 75–90. https://doi.org/10.1037//0033-295X.109.1.75

Gordon, P. C., Lowder, M. W., & Hoedemaker, R. S. (2016). Reading in normally aging adults. *Cognition, language and aging*, 165-191.

Graesser, A. C., Hoffman, N. L., & Clark, L. F. (1980). Structural components of reading time. *Journal of Verbal Learning and Verbal Behavior*, *19*(2), 135–151. https://doi.org/10.1016/S0022-5371(80)90132-2

Grillo, N., Alexiadou, A., Gehrke, B., Hirsch, N., Paolazzi, C., & Santi, A. (2019). Processing unambiguous verbal passives in German. *Journal of Linguistics*, 523–562.

Haberlandt, K. (1994). Methods in reading research. *Handbook of Psycholinguistics*, 1–31.

Haberlandt, K. F., Graesser, A. C., Schneider, N. J., & Kiely, J. (1986). Effects of task and new arguments on word reading times. *Journal of Memory and Language*, *25*(3), 314–322. https://doi.org/10.1016/0749-596X(86)90004-5

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

Haeuser, K. I., Baum, S., & Titone, D. (2021). Effects of aging and noncanonical form presentation on idiom processing : Evidence from eye tracking. *Applied Psycholinguistics*, *42*, 101–127. https://doi.org/10.1017/S0142716420000612

Hagoort, P. (2003). Interplay between syntax and semantics during sentence comprehension: ERP effects of combining syntactic and semantic violations. *Journal of Cognitive Neuroscience*, *15*(6), 883–899. https://doi.org/10.1162/089892903322370807

Hahn, G. A., Messias, A., MacKeben, M., Dietz, K., Horwath, K., Hyvärinen, L., Leinonen, M., & Trauzettel-Klosinski, S. (2009). Parafoveal letter recognition at reduced contrast in normal aging and in patients with risk factors for AMD. *Graefe's Archive for Clinical and Experimental Ophthalmology*, *247*(1), 43–51. https://doi.org/10.1007/s00417-008-0919-z

Hannon, B. (2015). Conscious and unconscious detection of semantic anomalies. *Quarterly Journal of Experimental Psychology*, *68*(7), 1404–1425. https://doi.org/10.1080/17470218.2014.982138

Hannon, B., & Daneman, M. (2001). Susceptibility to semantic illusions: An individual-differences perspective. *Memory and Cognition*, *29*(3), 449–461. https://doi.org/10.3758/BF03196396

Hannon, B., & Daneman, M. (2004). Shallow semantic processing of text: An individual-differences account. *Discourse Processes*, *37*(3), 187–204. https://doi.org/10.1207/s15326950dp3703_1

Hansen, D. W., & Ji, Q. (2010). In the Eye of the Beholder: A Survey of Models for Eyes and Gaze. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *32*(3), 478–500. https://doi.org/10.1109/TPAMI.2009.30

Hardy, S. M., Messenger, K., & Maylor, E. A. (2017). Aging and syntactic representations: Evidence of preserved syntactic priming and lexical boost. *Psychology and Aging*, *32*(6), 588–596. https://doi.org/10.1037/pag0000180

Harrison, T. M., Weintraub, S., Mesulam, M. M., & Rogalski, E. (2012). Superior memory and higher cortical volumes in unusually successful cognitive aging. *Journal of the*

International Neuropsychological Society, *18*(6), 1081-1085.

Hermena, E. W., Bouamama, S., Liversedge, S. P., & Drieghe, D. (2021). Does diacritics-based lexical disambiguation modulate word frequency, length, and predictability effects? An eye-movements investigation of processing Arabic diacritics. *Plos one*, *16*(11), e0259987.

Hobbs, J. R. (1985). On the coherence and structure of discourse.

Howman, H. E., & Filik, R. (2020). The role of emoticons in sarcasm comprehension in younger and older adults: Evidence from an eye-tracking experiment. *Quarterly Journal of Experimental Psychology*, 174702182092280. https://doi.org/10.1177/1747021820922804

Huang, Y., & Ferreira, F. (2021). What causes lingering misinterpretations of garden-path sentences: Incorrect syntactic representations or fallible memory processes? *Journal of Memory and Language, 121*(September), 104288. https://doi.org/10.1016/j.jml.2021.104288

Huey, E. B. (1908). The psychology and pedagogy of reading. New York: Macmillan

Ioup, G. (1975). *Some universals for quantifier scope. In J. Kimball (Ed.), Syntax and semantics* (p. Vol 4 pg. 37-48). New York: Academic Press.

Irving, E. L., Steinbach, M. J., Lillakas, L., Babu, R. J., & Hutchings, N. (2006). Horizontal saccade dynamics across the human life span. *Investigative Ophthalmology and Visual Science*, *47*(6), 2478–2484. https://doi.org/10.1167/iovs.05-1311

Jacob, G., & Felser, C. (2016). Reanalysis and semantic persistence in native and non-native garden-path recovery. *Quarterly Journal of Experimental Psychology*, *69*(5), 907–925. https://doi.org/10.1080/17470218.2014.984231

Janse, E., Werff, M. Van Der, & Quené, H. (2007). Listening to fast speech: aging and sentence context. *Proceedings of the 16th …*, *August*, 681–684.

http://www.researchgate.net/publication/46693503_Listening_to_fast_speech_aging_and_sentence_context/file/e0b49529c3bd01def3.pdf

Jegerski, J. (2014). *Self-paced reading*. *January*.

Jiang, X., & Zhou, X. (2020). An alternative structure rescues failed semantics? Strong global expectancy reduces local-mismatch N400 in Chinese flexible structures. *Neuropsychologia*, *140*(February), 107380. https://doi.org/10.1016/j.neuropsychologia.2020.107380

Jones, P. R. (2019). A note on detecting statistical outliers in psychophysical data. *Attention, Perception, and Psychophysics*, *81*(5), 1189–1196. https://doi.org/10.3758/s13414-019-01726-3

Jordan, T. R., Mcgowan, V. A., & Paterson, K. B. (2014). Reading with filtered fixations: adult age differences in the effectiveness of low-level properties of text within central vision. *Psychology and Aging*, *29*, 229–235. www.journal.uta45jakarta.ac.id

Judd, C. M., Westfall, J., & Kenny, D. A. (2016). Experiments with more than one random factor: Designs, analytic models, and statistical power. *Annual Review of Psychology*.Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, *87*(4), 329–354. https://doi.org/10.1037/0033-295X.87.4.329

Just, M. A., Carpenter, P. A., & Keller, T. A. (1996). The capacity theory of comprehension: New frontiers of evidence and arguments. *Psychological Review*, *103*(4), 773–780. https://doi.org/10.1037//0033-295x.103.4.773

Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General*, *111*(2), 228–238. https://doi.org/10.1037/0096-3445.111.2.228

Karimi, H., & Ferreira, F. (2016). Good-enough linguistic representations and online cognitive equilibrium in language processing. *Quarterly Journal of Experimental Psychology, 69*(5), 1013–1040. https://doi.org/10.1080/17470218.2015.1053951

Kaschak, M. P., & Glenberg, A. M. (2004). This construction needs learned. *Journal of Experimental Psychology: General*, *133*(3), 450–467. https://doi.org/10.1037/0096-3445.133.3.450

Kemper, S., Crow, A., & Kemtes, K. (2004). Eye-fixation patterns of high- and low-span young and older adults: down the garden path and back again. *Psychology and Aging*, *19*(1), 157–170. https://doi.org/10.1037/0882-7974.19.1.157

Kemper, S., & Liu, C. (2007). Eye movements of young and older adults during reading. *Psychology and Aging*, *2*, 84–93. https://doi.org/10.1038/jid.2014.371

Kemper, S., & McDowd, J. (2006). Eye movements of young and older adults while reading with distraction. *Psychology and Aging*, *21*, 32–39. https://doi.org/10.1038/jid.2014.371

Kemper, S., McDowd, J., Metcalf, K., & Liu, C. J. (2008). Young and older adults' reading of distracters. *Educational Gerontology*, *34*(6), 489–502. https://doi.org/10.1080/03601270701835858

Kemtes, K. A., & Kemper, S. (1997). Younger and older adults' on-line processing of syntactically ambiguous sentences. *Psychology and Aging*, *12*(2), 362–371. https://doi.org/10.1037/0882-7974.12.2.362

Kemtes, K. A., & Kemper, S. (1999a). Aging and resolution of quantifier scope effects. *Journals of Gerontology - Series B Psychological Sciences and Social Sciences*, *54*(6). https://doi.org/10.1093/geronb/54B.6.P350

Kemtes, K. A., & Kemper, S. (1999b). Aging and Resolution of Quantifier Scope Effects. *Journal of Gerontology: Psychological Sciences*, *54*(6), 35–36.

Keuleers, E., Stevens, M., Mandera, P., & Brysbaert, M. (2015). Word knowledge in the crowd: Measuring vocabulary size and word prevalence in a massive online experiment.

Quarterly Journal of Experimental Psychology, *68*(8), 1665–1692. https://doi.org/10.1080/17470218.2015.1022560

Kizach, J. (2015). Animacy and the ordering of postverbal prepositional phrases in Danish. *Acta Linguistica Hafniensia*, *47*(2), 199–219. https://doi.org/10.1080/03740463.2015.1089087

Kliegl, R., Dambacher, M., Dimigen, O., Jacobs, A. M., & Sommer, W. (2012). Eye movements and brain electric potentials during reading. *Psychological Research*, *76*(2), 145–158. https://doi.org/10.1007/s00426-011-0376-x

Kliegl, R., Grabner, E., Rolfs, M., & Engbert, R. (2004). Length, frequency, and predictability effects of words on eye movements in reading. *European Journal of Cognitive Psychology*, *16*(1–2), 262–284. https://doi.org/10.1080/09541440340000213

Koornneef, A., & Mulders, I. (2017). Can we ‘read’ the eye-movement patterns of readers? Unraveling the relationship between reading profiles and processing strategies. *Journal of Psycholinguistic Research*, *46*(1), 39–56. https://doi.org/10.1007/s10936-016-9418-2

Kurtzman, H. S., & Macdonald, M. C. (1993). Resolution of quantifier scope ambiguities. *Cognition*, *48*, 243–279.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event related brain potential (ERP). *Annu Rev Psychol.*, *62*, 621–647. https://doi.org/10.1146/annurev.psych.093008.131123.Thirty

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*(4427), 203–205.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in linear mixed effects models . *Journal of Statistical Software*, *82*(13). https://doi.org/10.18637/jss.v082.i13

Lakoff, G. (1971). *On generative semantics* (pp. 232–296). Cambridge: Cambridge

University Press.

Laver, G. D., & Burke, D. M. (1993). Why do semantic priming effects increase in old age? A meta-Analysis. *Psychology and Aging*, *8*(1), 34–43. https://doi.org/10.1037/0882-7974.8.1.34

Legge, G. E., Ross, J. A., Isenberg, L. M., & LaMay, J. M. (1992). Psychophysics of reading: Clinical predictors of low-vision reading speed. *Investigative Ophthalmology and Visual Science*, *33*(3), 677–687.

Levy, R. (2008). A noisy-channel model of rational human sentence comprehension under uncertain input. *EMNLP 2008 - 2008 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference: A Meeting of SIGDAT, a Special Interest Group of the ACL, October*, 234–243. https://doi.org/10.3115/1613715.1613749

Lim, J. H., & Christianson, K. (2013). Second language sentence processing in reading for comprehension and translation. *Bilingualism*, *16*(3), 518–537. https://doi.org/10.1017/S1366728912000351

Liu, R., Patel, B. N., & Kwon, M. Y. (2017). Age-related changes in crowding and reading speed. *Scientific Reports*, *7*(1), 1–10. https://doi.org/10.1038/s41598-017-08652-0

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

Liversedge, S. P., Paterson, K. B., Pickering, M. J., Liversedge, S. P., Paterson, K. B., & Pickering, M. J. (1998). Eye Movements and Measures of Reading Time. In *In Eye guidance in reading and scene perception. Elsevier Science Ltd* (pp. 55–75).

Lo, S., & Andrews, S. (2015). To transform or not to transform: using generalized linear mixed models to analyse reaction time data. *Frontiers in Psychology*, *6*, 1–16. https://doi.org/10.3389/fpsyg.2015.01171

Locker, L. J., Hoffman, L., & Bovaird, J. A. (2007). On the use of multilevel modeling as an alternative to items analysis in psycholinguistic research. *Behavior Research Methods*, *39*(4), 723–730.

López Zunini, R. A., Baart, M., Samuel, A. G., & Armstrong, B. C. (2020). Lexical access versus lexical decision processes for auditory, visual, and audiovisual items: Insights from behavioral and neural measures. *Neuropsychologia*, *137*, 1–52. https://doi.org/10.1016/j.neuropsychologia.2019.107305

MacDonald, M. C. (1994). Probabilistic Constraints and Syntactic Ambiguity Resolution. *Language and Cognitive Processes*, *9*(2), 157–201. https://doi.org/10.1080/01690969408402115

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). Lexical nature of syntactic ambiguity resolution. *Psychological Review*, *101*(4), 676–703. https://doi.org/10.1037//0033-295x.101.4.676

Malyutina, S., & den Ouden, D. B. (2016). What is it that lingers? Garden-path (mis)interpretations in younger and older adults. *Quarterly Journal of Experimental Psychology*, *69*(5), 880–906. https://doi.org/10.1080/17470218.2015.1045530

McConkie, G. W., & Rayner, K. (1975). The span of the effective stimulus during a fixation in reading. *Perception & Psychophysics*, *17*(6), 578–586. https://doi.org/10.3758/BF03203972

McGowan, V. A., & Reichle, E. D. (2018). The “risky” reading strategy revisited: New simulations using E-Z reader. *Quarterly Journal of Experimental Psychology*, *71*(1 Special Issue), 179–189. https://doi.org/10.1080/17470218.2017.1307424

McGowan, V. A., White, S. J., Jordan, T. R., & Paterson, K. B. (2014). Aging and the use of interword spaces during reading: Evidence from eye movements. *Psychonomic Bulletin and Review*, *21*(3), 740–747. https://doi.org/10.3758/s13423-013-0527-8

McGowan, V. A., White, S. J., & Paterson, K. B. (2014). Aging and the use of interword spaces during reading: Evidence from eye movements. *Psychonomic Bulletin & Review*, *21*(3), 740-747.. 1–31.

McGowan, V. A., White, S. J., & Paterson, K. B. (2015). The effects of interword spacing on the eye movements of young and older readers. *Journal of Cognitive Psychology*, *27*(5), 609–621.

Mecklinger, A., Schriefers, H., Steinhauer, K., & Friederici, A. D. (1995). Processing relative clauses varying on syntactic and semantic dimensions: An analysis with event-related potentials. *Memory & Cognition*, *23*(4), 477–494. https://doi.org/10.3758/BF03197249

Meng, M., & Bader, M. (2021). Does comprehension (sometimes) go wrong for noncanonical sentences ? *Quarterly Journal of Experimental Psychology*, *74*(1), 1–28. https://doi.org/10.1177/1747021820947940

Metzner, P., von der Malsburg, T., Vasishth, S., & Rösler, F. (2017). The importance of reading naturally: evidence from combined recordings of eye movements and electric brain potentials. *Cognitive Science*, *41*, 1232–1263. https://doi.org/10.1111/cogs.12384

Miller, L. M. S., & Stine-Morrow, E. A. L. (1998). Aging and the effects of knowledge on on-line reading strategies. *Journals of Gerontology - Psychological Sciences and Social Sciences*, *53*(4), 223–233. https://doi.org/10.1093/geronb/53B.4.P223

Mitchell, D. C., & Green, D. W. (1978). The effects of context and content on immediate processing in reading. *Quarterly Journal of Experimental Psychology*, *30*(4), 609-636.

Moreno, J. D., León, J. A., Arnal, L. A. M., & Botella, J. (2019). Age differences in eye movements during deading: Degenerative problems or compensatory Strategy?: A meta-analysis. *European Psychologist*, *24*(4), 297–311. https://doi.org/10.1027/1016-9040/a000344

Morrow, D. G., Stine-Morrow, E. A. L., Von Leirer, O., Andrassy, J. M., & Kahn, J. (1997). The role of reader age and focus of attention in creating situation models from narratives. *Journals of Gerontology - Series B Psychological Sciences and Social Sciences*, *52*(2), 73–80. https://doi.org/10.1093/geronb/52B.2.P73

Mousavi, S., & Gigerenzer, G. (2014). Risk, uncertainty, and heuristics. *Journal of Business Research*, *67*(8), 1671–1678. https://doi.org/10.1016/j.jbusres.2014.02.013

Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J. L., & Chertkow, H. (2005). The montreal cognitive assessment, MoCA: A brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society*, *53*(4), 695–699. https://doi.org/10.1111/j.1532-5415.2005.53221.x

Nicholls, V. I., Wiener, J. M., Meso, A. I., & Miellet, S. (2022). The relative contribution of executive functions and aging on attentional control during road crossing. *Frontiers in psychology*, 2432.

Nieuwenhuis, S., Ridderinkhof, K. R., De Jong, R., Kok, A., & Van Der Molen, M. W. (2000). Inhibitory inefficiency and failures of intention activation: Age-related decline in the control of saccadic eye movements. *Psychology and Aging*, *15*(4), 635–647. https://doi.org/10.1037/0882-7974.15.4.635

O’Brien, E. J., & Cook, A. E. (2016a). Coherence threshold and the continuity of processing: The RI-Val model of comprehension. *Discourse Processes*, *53*(5–6), 326–338. https://doi.org/10.1080/0163853X.2015.1123341

O’Brien, E. J., & Cook, A. E. (2016b). Separating the activation, integration, and validation Components of reading. In *Psychology of Learning and Motivation - Advances in Research and Theory* (Vol. 65). Elsevier Ltd. https://doi.org/10.1016/bs.plm.2016.03.004

Obler, L. K., Fein, D., Nicholas, M., & Albert, M. L. (1991). Auditory comprehension and aging: Decline in syntactic processing. *Applied Psycholinguistics*, *12*(4), 433–452. https://doi.org/10.1017/S0142716400005865

O'Regan, J. K. (1992). Optimal viewing position in words and the strategy-tactics theory of eye movements in reading. In *Eye movements and visual cognition* (pp. 333-354). Springer, New York, NY.

Osborne, J. W., & Overbay, A. (2004). The power of outliers (and why researchers should ALWAYS check for them). *Practical Assessment, Research and Evaluation*, *9*(6).

Osterhout, L. E. E., & Holcomb, P. J. (1993). Journal of memory and language 2002. *Language and Cognitive Processes*, *8*(4), 439–483.

Owsley, C. (2011). Aging and vision. *Vision Research*, *51*(13), 1610–1622. https://doi.org/10.1016/j.visres.2010.10.020

Owsley, C., Sekuler, R., & Siemsen, D. (1983). Contrast sensitivity throughout adulthood. *Vision Research*, *23*(7), 689–699. https://doi.org/10.1016/0042-6989(83)90210-9

Paape, D. (n.d.). *The depth charge illusion, linear order, and "negated worlds" – Competence, performance, or both?*

Paape, D., Vasishth, S., & von der Malsburg, T. (2020). Quadruplex negatio invertit? The on-line processing of depth charge sentences. *Journal of Semantics*, *37*(4), 509–555. https://doi.org/10.1093/jos/ffaa009

Park, D. C., Lautenschlager, G., Smith, A. D., Earles, J. L., Frieske, D., Zwahr, M., & Gaines, C. L. (1996). Mediators of long-term memory performance across the life span. *Psychology and Aging*, *11*(4), 621–637. https://doi.org/10.1037/0882-7974.11.4.621

Paterson, K. B., Filik, R., & Liversedge, S. P. (2008). Competition during the processing of quantifier scope ambiguities: Evidence from eye movements during reading. *Quarterly Journal of Experimental Psychology*. https://doi.org/10.1080/17470210701255317

Paterson, K. B., McGowan, V. A., & Jordan, T. R. (2013a). Aging and the control of binocular fixations during reading. *Psychology and Aging*, *28*(3), 789–795. https://doi.org/10.1037/a0033328

Paterson, K. B., McGowan, V. A., & Jordan, T. R. (2013b). Filtered text reveals adult age differences in reading: Evidence from eye movements. *Psychology and Aging*, *28*(2), 352–364. https://doi.org/10.1037/a0030350

Paterson, K. B., McGowan, V. A., Warrington, K. L., Li, L., Li, S., Xie, F., Chang, M., Zhao, S., Pagán, A., White, S. J., & Wang, J. (2020). Effects of normative aging on eye movements during reading. *Vision (Switzerland)*, *4*(1), 1–18. https://doi.org/10.3390/vision4010007

Patson, N. D., & Warren, T. (2010). Evidence for Distributivity Effects in Comprehension. *Journal of Experimental Psychology: Learning Memory and Cognition*, *36*(3), 782–789. https://doi.org/10.1037/a0018783

Payne, B. R., & Federmeier, K. D. (2019). Contextual constraints on lexico-semantic processing in aging: Evidence from single-word event-related brain potentials. *Brain Research*, *1687*, 117–128. https://doi.org/10.1016/j.brainres.2018.02.021.Contextual

Payne, B. R., Grison, S., Gao, X., Christianson, K., Morrow, D. G., & Stine-Morrow, E. A. L. (2014). Aging and individual differences in binding during sentence understanding: Evidence from temporary and global syntactic attachment ambiguities. *Cognition*, *130*(2), 157–173. https://doi.org/10.1016/j.cognition.2013.10.005

Payne, B. R., & Stine-Morrow, E. A. (2012a). Aging, parafoveal preview, and semantic integration in sentence processing: Testing the cognitive workload of wrapup. *Psychology and Aging*, *27*(3), 638–649. https://doi.org/10.1037/a0026540.Aging

Payne, B. R., & Stine-Morrow, E. A. L. (2012b). Aging, parafoveal Preview, and semantic integration in sentence processing: testing the cognitive workload of wrap- up. *Psychology and Aging*, *27*(3), 638–649. https://doi.org/10.1037/a0026540.Aging

Payne, B. R., & Stine-Morrow, E. A. L. (2014). Adult age differences in wrap-up during sentence comprehension: Evidence from ex-gaussian distributional analyses of reading time. *Psychology and Aging*, *29*(2), 213–228. https://doi.org/10.1037/a0036282

Pelli, D. G., & Robson, J. G. (1988). The design of a new letter chart for measuring contrast sensitivity. In *Clinical Vision Sciences*.

Peltsch, A., Hemraj, A., Garcia, A., & Munoz, D. P. (2011). Age-related trends in saccade characteristics among the elderly. *Neurobiology of Aging*, *32*(4), 669–679. https://doi.org/10.1016/j.neurobiolaging.2009.04.001

Pickering, M. J., & Traxler, M. J. (1998). Plausibility and recovery from garden paths: An eye-tracking study. *Journal of Experimental Psychology: Learning Memory and Cognition*, *24*(4), 940–961. https://doi.org/10.1037/0278-7393.24.4.940

Płużyczka, M. (2018). The first hundred years: a history of eye tracking as a research method. *Applied Linguistics Papers*, *4*(25), 101–116. https://doi.org/10.32612/uw.25449354.2018.4.pp.101-116

Poesio, M., Sturt, P., Artstein, R., & Filik, R. (2006). Underspecification and anaphora: Theoretical issues and preliminary evidence. *Discourse Processes*, *42*(2), 157–175. https://doi.org/10.1207/s15326950dp4202_4

Pollatsek, A., Juhasz, B. J., Reichle, E. D., & Rayner, K. (2008). Immediate and delayed effects of word frequency and word length on eye movements in reading: a reversed delayed effect of word length. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(3), 726–750. https://doi.org/10.1037/0096-1523.34.3.726.Immediate

Pollatsek, A., Reichle, E. D., & Rayner, K. (2006). Tests of the E-Z Reader model: Exploring the interface between cognition and eye-movement control. *Cognitive Psychology*, *52*(1), 1–56. https://doi.org/10.1016/j.cogpsych.2005.06.001

Pollet, T. V., & Van Der Meij, L. (2017). To remove or not to remove: the impact of outlier handling on significance testing in testosterone data. *Adaptive Human Behavior and*

Physiology, *3*(1), 43-60.

Poole, A., & Ball, L. J. (2005). Eye tracking in human-computer interaction and usability research: Current status and future prospects. *United Kingdom: Psychology Department, Lancaster University.*, 211–219. https://doi.org/10.4018/978-1-59140-562-7

Price, C. J. (2012). A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *NeuroImage*, *62*(2), 816–847. https://doi.org/10.1016/j.neuroimage.2012.04.062

Qian, Z., Garnsey, S., & Christianson, K. (2018). A comparison of online and offline measures of good-enough processing in garden-path sentences. *Language, Cognition and Neuroscience*. https://doi.org/10.1080/23273798.2017.1379606

Quene, H., & Van Den Bergh, H. (2004). On multi-level modeling of data from repeated measures designs : a tutorial. *Speech Communication*, *43*(1–2), 103–121. https://doi.org/10.1016/j.specom.2004.02.004

R Development Core Team (2019). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.

Raajmakers, J. G. W. (2003). A Further Look at the “ Language-as-Fixed-Effect Fallacy .” *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *57*(3), 141–151.

Radner, W. (2017). Reading charts in ophthalmology. *Graefe’s Archive for Clinical and Experimental Ophthalmology*, *255*(8), 1465–1482. https://doi.org/10.1007/s00417-017-3659-0

Radner, W., Obermayer, W., Richter-Mueksch, S., Willinger, U., Velikay-Parel, M., & Eisenwort, B. (2002). The validity and reliability of short German sentences for measuring reading speed. *Graefe’s Archive for Clinical and Experimental Ophthalmology*, *240*(6), 461–467. https://doi.org/10.1007/s00417-002-0443-5

Radvansky, G. A., Copeland, D. E., & Zwaan, R. A. (2003). Aging and functional spatial relations in comprehension and memory. *Psychology and Aging*, *18*(1), 161–165. https://doi.org/10.1037/0882-7974.18.1.161

Radvansky, G. A., & Dijkstra, K. (2003). Aging and situation model processing. *Aging, Neuropsychology and Cognition*, *10*(2), 158–166. https://doi.org/10.3758/BF03193088

Radvansky, G. A., Gerard, L. D., Zacks, R. T., & Hasher, L. (1990). Younger and older adults' use of mental models as representations for text materials. *Psychology and Aging*, *5*(2), 209–214. https://doi.org/10.1037/0882-7974.5.2.209

Radvansky, G. A., Zwaan, R. A., Curiel, J. M., & Copeland, D. E. (2001). Situation models and aging. *Psychology and Aging*, *16*(1), 145–160.

Raffray, C. N., & Pickering, M. J. (2010). How do people construct logical form during language comprehension? *Psychological Science*, *21*(8), 1090–1097. https://doi.org/10.1177/0956797610375446

Rayner, K. (1978). Eye movements in reading and information processing. *Psychological Bulletin*, *85*(3), 618–660. https://doi.org/10.1037/0033-2909.85.3.618

Rayner, K. (1985). The role of eye movements in learning to read and reading disability. *Remedial and Special Education*, *6*(6), 53-60.

Rayner, K. (1998). Cognitive control of eye movements. *Psychological Bulletin*, *124*(3), 372–422. http://content.apa.org/journals/bul/124/3/372%5Cnpapers2://publication/uuid/EE315426-1E21-4F51-9908-80F30C62CDD4%0Ahttp://doi.apa.org/getdoi.cfm?doi=10.1037/0096-1523.22.5.1188

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. In *Quarterly Journal of Experimental Psychology* (Vol. 62, Issue 8). https://doi.org/10.1080/17470210902816461

Rayner, K., Castelhano, M. S., & Yang, J. (2009). Eye movements and the perceptual span in older and younger readers. *Psychology and Aging*, *24*(3), 755–760. https://doi.org/10.1037/a0014300

Rayner, K., Castelhano, M. S., & Yang, J. (2010). Preview benefit during eye fixations in reading for older and younger readers. *Psychology and Aging*, *25*(3), 714–718. https://doi.org/10.1037/a0019199

Rayner, K., Chace, K. H., Slattery, T. J., & Ashby, J. (2006). Eye movements as reflections of comprehension processes in reading. *Scientific Studies of Reading*, *10*(3), 241–255. http://www.tandfonline.com/doi/full/10.1207/s1532799xssr1003_3%5Cnhttp://www.tandfonline.com/doi/pdf/10.1207/s1532799xssr1003_3

Rayner, K., Fischer, M. H., & Pollatsek, A. (1998). Unspaced text interferes with both word identification and eye movement control. *Vision Research*, *38*(8), 1129–1144. https://doi.org/10.1016/S0042-6989(97)00274-5

Rayner, K., Kambe, G., & Duffy, S. A. (2000). The effect of clause wrap-up on eye movements during reading. *The Quarterly Journal of Experimental Psychology*, *53A*(4), 1061–1080. https://doi.org/10.1016/j.apergo.2021.103523

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

Rayner, K., Reichle, E. D., & Pollatsek, A. (1998). Eye movement control in reading: An overview and model. *Eye Guidance in Reading and Scene Perception*, 243–268. http://books.google.com/books?hl=en&lr=&id=NmXqAV1vAF4C&pgis=1

Rayner, K., Reichle, E. D., Stroud, M. J., Williams, C. C., & Pollatsek, A. (2006). The effect of word frequency, word predictability, and font difficulty on the eye movements of young and older readers. *Psychology and Aging*, *21*(3), 448–465. https://doi.org/10.1037/0882-7974.21.3.448

Rayner, K., Well, A. D., Pollatsek, A., & Bertera, J. H. (1982). The availabiltiy of useful information to the right of fixation in reading. *Perception and Psychophysics*, *31*(6), 537–550.

Rayner, K., Yang, J., Castelhano, M. S., & Liversedge, S. P. (2011). Eye movements of older and younger readers when reading disappearing text. *Psychology and Aging*, *26*(1), 214–223. https://doi.org/10.1037/a0021279

Rayner, K., Yang, J., Schuett, S., & Slattery, T. J. (2013). Eye movements of older and younger readers when reading unspaced text. *Experimental Psychology*, *60*(5), 354–361. https://doi.org/10.1027/1618-3169/a000207

Reder, L. M., & Kusbit, G. W. (1991). Locus of the Moses Illusion: Imperfect encoding, retrieval, or match? *Journal of Memory and Language*, *30*, 385–406.

Reichle, E. D., Liversedge, S. P., Drieghe, D., Blythe, H. I., Joseph, H. S. S. L., White, S. J., & Rayner, K. (2013). Using E-Z Reader to examine the concurrent development of eye-movement control and reading skill. *Developmental Review*, *33*(2), 110–149. https://doi.org/10.1016/j.dr.2013.03.001

Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review*, *105*(1), 125–157. https://doi.org/10.1037/0033-295X.105.1.125

Richmond, L. L., Brackins, T., & Rajaram, S. (2022). Episodic Memory Performance Modifies the Strength of the Age–Brain Structure Relationship. *International Journal of Environmental Research and Public Health*, *19*(7), 4364.

Risse, S., & Kliegl, R. (2011). Adult age differences in the perceptual span during reading. *Psychology and Aging*, *26*(2), 451–460. https://doi.org/10.1037/a0021616

Sachs, J. S. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. *Perception & Psychophysics*, *2*(9), 437-442.

Salthouse, T. (2012). Consequences of age-related cognitive declines. *Annual Review of Psychology, 63*, 201–226. https://doi.org/10.1146/annurev-psych-120710-100328

Salthouse, T. A. (1990). Influence of experience on age differences in cognitive functioning. *Human Factors*, *32*(5), 551–569. https://doi.org/10.1177/001872089003200505

Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review*, *103*(3), 403–428. https://doi.org/10.1037/0033-295X.103.3.403

Salthouse, T. A. (2009). When does age-related cognitive decline begin? *Neurobiology of Aging*, *30*(4), 507–514. https://doi.org/10.1016/j.neurobiolaging.2008.09.023

Salthouse, T. A., & Babcock, R. L. (1991). Decomposing adult age differences in working memory. *Developmental Psychology*, *27*(5), 763–776. https://doi.org/10.1037/0012-1649.27.5.763

Salthouse, T. A., & Meinz, E. J. (1995). Aging, inhibition, working memory, and speed. *Journals of Gerontology - Series B Psychological Sciences and Social Sciences*, *50 Series*(6), P297–P306. https://doi.org/10.1093/geronb/50B.6.P297

Sanford, A. J., & Graesser, A. C. (2006). Shallow discourse processing. *Discourse Processes*, *42*(2), 99–108. https://doi.org/10.1207/s15326950dp4202

Sanford, A. J., Leuthold, H., Bohan, J., & Sanford, A. J. S. (2011). Anomalies at the borderline of awareness: An ERP study. *Journal of Cognitive Neuroscience*, *23*(3), 514–523. https://doi.org/10.1162/jocn.2009.21370

Sanford, A. J. S., Sanford, A. J., Filik, R., & Molle, J. (2005). Depth of lexical-semantic processing and sentential load. *Journal of Memory and Language*, *53*(3), 378–396. https://doi.org/10.1016/j.jml.2005.05.004

Sanford, A. J., & Sturt, P. (2002). Depth of processing in language comprehension: Not noticing the evidence. *Trends in Cognitive Sciences*, *6*(9), 382–386. https://doi.org/10.1016/S1364-6613(02)01958-7

Schaie, K. W. (1994). The course of adult intellectual development. *American Psychologist*, *49*(4), 304–313. https://doi.org/10.1037//0003-066x.49.4.304

Scheffers, M. K., Humphrey, D. G., Stanny, R. R., Kramer, A. F., & Coles, M. G. (1999). Error-related processing during a period of extended wakefulness. *Psychophysiology*, *36*(2), 149-157.

Schmalhofer, F., & Glavanov, D. (1986). Three components of understanding a programmer's manual: Verbatim, propositional, and situational representations. *Journal of Memory and Language*, *25*(3), 279–294. https://doi.org/10.1016/0749-596X(86)90002-1

Schotter, E. R., Tran, R., & Rayner, K. (2014). Don't believe what you read (Only Once): Comprehension is supported by regressions during reading. *Psychological Science*, *25*(6), 1218–1226. https://doi.org/10.1177/0956797614531148

Schroeder, S. (2011). What readers have and do: effects of students' verbal ability and reading time components on comprehension with and without text availability. *Journal of Educational Psychology*, *103*(4), 877–896. https://doi.org/10.1037/a0023731

Sereno, S. C., & Rayner, K. (2003). Measuring word recognition in reading: Eye movements and event-related potentials. *Trends in Cognitive Sciences*, *7*(11), 489–493. https://doi.org/10.1016/j.tics.2003.09.010

Seyednozadi, Z., Pishghadam, R., & Pishghadam, M. (2021). Functional role of the N400 and P600 in language-related ERP studies with respect to semantic anomalies : An overview. *Archives of Neuropsychiatry*.

Shake, M. C., & Stine-Morrow, E. A. L. (2011). Age differences in resolving anaphoric expressions during reading. *Aging, Neuropsychology, and Cognition*, *18*(6), 678–707. https://doi.org/10.1080/13825585.2011.607228

Sheldon, S., Pichora-Fuller, M. K., & Schneider, B. A. (2008). Priming and sentence context support listening to noise-vocoded speech by younger and older adults. *The Journal of the Acoustical Society of America*, *123*(1), 489–499. https://doi.org/10.1121/1.2783762

Sikos, L., Duffield, C. J., & Kim, A. E. (2016). Grammatical predictions reveal influences of semantic attraction in online sentence comprehension: evidence from speeded forced-choice sentence continuations. *Language, Cognition and Neuroscience*, *31*(8), 1055–1073. https://doi.org/10.1080/23273798.2016.1186808

Singh, H., & Singh, J. (2012). Human eye tracking and related issues: A Review. *International Journal of Scientific and Research Publications*, *2*(1), 2250–3153. www.ijsrp.org

Slattery, T. J. (2009). Word misperception, the neighbor frequency effect, and the role of sentence context: Evidence from eye movements. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(6), 1969–1975. https://doi.org/10.1037/a0016894

Slattery, T. J., Sturt, P., Christianson, K., Yoshida, M., & Ferreira, F. (2013). Lingering misinterpretations of garden path sentences arise from competing syntactic representations. *Journal of Memory and Language*, *69*(2), 104–120. https://doi.org/10.1016/j.jml.2013.04.001

Starr, M. S., & Rayner, K. (2001). Eye movements during reading: some current controversies. *Trends in Cognitive Sciences*, *5*(4), 156–163. https://doi.org/10.4324/9781315785004-18

Staub, A., & Benatar, A. (2013). Individual differences in fixation duration distributions in reading. *Psychonomic Bulletin and Review*, *20*(6), 1304–1311. https://doi.org/10.3758/s13423-013-0444-x

Staub, A., & Rayner, K. (2007). Eye movements and on-line comprehension processes. In *The Oxford handbook of psycholinguistics* (pp. 327–342).

https://doi.org/10.1080/01690968908406362

Stine-Morrow, E. A. L., Loveless, M. K., & Soederberg, L. M. (1996). Resource allocation in on-line reading by younger and older adults. *Psychology and Aging*, *11*(3), 475–486. https://doi.org/10.1037/0882-7974.11.3.475

Stine-Morrow, E. A. L., Milinder, L., Pullara, O., & Herman, B. (2001). Patterns of resource allocation are reliable among younger and older readers. *Psychology and Aging*, *16*(1), 69–84. https://doi.org/10.1037/0882-7974.16.1.69

Stine-Morrow, E. A. L., Miller, L. M. S., & Hertzog, C. (2006). Aging and self-regulated language processing. *Psychological Bulletin*, *132*(4), 582–606. https://doi.org/10.1037/0033-2909.132.4.582

Stine-morrow, E. A. L., Shake, M. C., Miles, J. R., Lee, K., Gao, X., & Mcconkie, G. (2010). Pay now or pay later: aging and the role of boundary salience in self-regulation of conceptual integration in sentence processing. *Psychology and Aging*, *25*(1), 168–176. https://doi.org/10.1037/a0018127.Pay

Sturt, P. (2007). Semantic re-interpretation and garden path recovery. *Cognition*, *105*(2), 477–488. https://doi.org/10.1016/j.cognition.2006.10.009

Sturt, P., Sanford, A. J., Stewart, A., & Dawydiak, E. (2004). Linguistic focus and change detection: an application of the change-detection paradigm. *Psychonomic Bulletin & Review*, *11*(5), 882–888.

Tan, M., Foltz, A., & Engelhardt, P. E. (2020). Task Sensitivity in L2 English Speakers ' Syntactic Processing : Evidence for Good-Enough Processing in Self-Paced Reading. *Frontiers in Psychology*, *11*, 575847. https://doi.org/10.3389/fpsyg.2020.575847

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*(5217), 1632–1634. https://doi.org/10.1126/science.7777863

Todd, P. M., & Gigerenzer, G. (2000). Précis of Simple heuristics that make us smart. *Behavioral and Brain Sciences*, *23*(5), 727–780.

Traxler, M. J., Corina, D. P., Morford, J. P., Hafer, S., & Hoversten, L. J. (2014). Deaf readers' response to syntactic complexity: Evidence from self-paced reading. *Memory and Cognition*, *42*(1), 97–111. https://doi.org/10.3758/s13421-013-0346-1

Tunstall, S. L. (1998). The interpretation of quantifiers: semantics & processing. *(Doctoral Dissertation, University of Massachusetts at Amherst)*.

Turcan, A., Howman, H., & Filik, R. (2020). Examining the role of context in written sarcasm comprehension: Evidence from eye-tracking during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(10), 1966–1976.

Tversky, A., & Kahneman, D. (1989). Rational choice and the framing of decisions. *In Multiple Criteria Decision Making and Risk Analysis Using Microcomputers (Pp. 81-126). Springer, Berlin, Heidelberg.*

Van Berkum, J. J. A., Brown, C. M., & Hagoort, P. (1999). Early referential context effects in sentence processing: Evidence from event-related brain potentials. *Journal of Memory and Language*, *41*(2), 147–182. https://doi.org/10.1006/jmla.1999.2641

Van Dyke, J. A., Johns, C. L., & Kukona, A. (2014). Low working memory capacity is only spuriously related to poor reading comprehension. *Cognition*, *131*(3), 373–403. https://doi.org/10.1016/j.cognition.2014.01.007

Van Oostendorp, H., & De Mul, S. (1990). Moses beats Adam: A semantic relatedness effect on a semantic illusion. *Acta Psychologica*, *74*(1), 35–46. https://doi.org/10.1016/0001-6918(90)90033-C

Vasishth, S., Von Der Malsburg, T., & Engelmann, F. (2013). What eye movements can tell us about sentence comprehension. *Wiley Interdisciplinary Reviews: Cognitive Science*, *4*(2), 125–134. https://doi.org/10.1002/wcs.1209

Von der Malsburg, T., & Angele, B. (2017). False positives and other statistical errors in standard analyses of eye movements in reading. *Journal of memory and language*, *94*, 119-133.

Wang, J., Li, L., Li, S., Xie, F., Chang, M., Paterson, K. B., White, S. J., & McGowan, V. A. (2018). Adult age differences in eye movements during reading: The evidence from Chinese. *Journals of Gerontology - Series B Psychological Sciences and Social Sciences*, *73*(4), 584–593. https://doi.org/10.1093/geronb/gbw036

Warabi, T., Kase, M., & Kato, T. (1984). Effect of aging on the accuracy of visually guided saccadic eye movement. *Annals of Neurology*, *16*(4), 449–454. https://doi.org/10.1002/ana.410160405

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. https://doi.org/10.1016/j.cognition.2008.12.011

Warrington, K. L., Mcgowan, V. A., Paterson, K. B., White, S. J., (2018). Effects of aging , word frequency and text stimulus quality on reading across the adult lifespan : Evidence from eye movements. *Journal of Experimental Psychology : Learning , Memory , and Cognition*, *44*(11), 1714–1729.

Warrington, K. L., White, S. J., & Paterson, K. B. (2018). Ageing and the misperception of words: Evidence from eye movements during reading. *Quarterly Journal of Experimental Psychology*, *71*(1 Special Issue), 75–84. https://doi.org/10.1080/17470218.2016.1251471

Wason, P. C., & Reich, S. S. (1979). A verbal illusion. *Quarterly Journal of Experimental Psychology*, *31*(4), 591-597.

Waters, G., & Caplan, D. (2005). The relationship between age, processing speed, working memory capacity, and language comprehension. *Memory*, *13*(3–4), 403–413.

https://doi.org/10.1080/09658210344000459

Wechsler, D. (2008). Wechsler adult intelligence scale–Fourth Edition (WAIS–IV). *San Antonio, TX: NCS Pearson*, *22*(498), 816–827.

Weiss, A. F., Kretzschmar, F., Schlesewsky, M., Bornkessel-Schlesewsky, I., & Staub, A. (2018). Comprehension demands modulate re-reading, but not first-pass reading behavior. *Quarterly Journal of Experimental Psychology*, *71*(1 Special Issue), 198–210. https://doi.org/10.1080/17470218.2017.1307862

Whitford, V., & Titone, D. (2016). Eye movements and the perceptual span during first- and second-language sentence reading in bilingual older adults. *Psychology and Aging*, *31*(1), 58–70. https://doi.org/10.1037/a0039971

Whitford, V., & Titone, D. (2017). The effects of word frequency and word predictability during first- and second-language paragraph reading in bilingual older and younger adults. *Psychology and Aging, 32*(2), 158–177.

Whitford, V., & Titone, D. A. (2019). Lexical entrenchment and cross-language activation : Two sides of the same coin for bilingual reading across the adult lifespan. *Bilingualism: Language And Cognition*, *22*(1), 58–77. https://doi.org/10.1017/S1366728917000554

Wlotko, E. W., & Federmeier, K. D. (2012). So that's what you meant! Event-related potentials reveal multiple aspects of context use during construction of message level meaning. *NeuroImage*, *62*(1), 356–366. https://doi.org/10.1016/j.neuroimage.2012.04.054.So

Wlotko, E. W., Lee, C. L., & Federmeier, K. D. (2010). Language of the aging brain: Event-related potential studies of comprehension in older adults. *Linguistics and Language Compass*, *4*(8), 623–638. https://doi.org/10.1111/j.1749-818X.2010.00224.x

Wonnacott, E., Joseph, H. S. S. L., Adelman, J. S., & Nation, K. (2016). Is children's reading "good enough"? Links between online processing and comprehension as children read syntactically ambiguous sentences. *Quarterly Journal of Experimental Psychology*, *69*(5), 855–879. https://doi.org/10.1080/17470218.2015.1011176

Wotschack, C., & Kliegl, R. (2013). Reading strategy modulates parafoveal-on-foveal effects in sentence reading. *Quarterly Journal of Experimental Psychology*, *66*(3), 548–562. https://doi.org/10.1080/17470218.2011.625094

Yoon, J., Campanelli, L., Goral, M., Marton, K., Eichorn, N., & Obler, L. K. (2015). The effect of plausibility on sentence comprehension among older adults and its relation to cognitive functions. *Experimental Aging Research*, *41*(3), 272–302. https://doi.org/10.1080/0361073X.2015.1021646

Zang, C., Wang, Y., Bai, X., Yan, G., Drieghe, D., & Liversedge, S. P. (2016). The use of probabilistic lexicality cues for word segmentation in Chinese reading. *Quarterly Journal of Experimental Psychology*, *69*(3), 548–560. https://doi.org/10.1080/17470218.2015.1061030

Zhang, J., Warrington, K. L., Li, L., Pagán, A., Paterson, K. B., White, S. J., & Mcgowan, V. A. (2022). Are older adults more risky readers ? Evidence from meta-analysis. *Psychology and Aging*.

Zhou, P., & Christianson, K. (2016). I "hear" what you're "saying": Auditory perceptual simulation, reading speed, and reading comprehension. *Quarterly Journal of Experimental Psychology*, *69*(5), 972–995. https://doi.org/10.1080/17470218.2015.1018282

Appendices

Appendix A: Full list of experimental stimuli for Experiment 1 and 3.

Experimental items for Experiment 1

1.	
Anomaly	Item
Anomalous	Few people ever walk down those secret corridors of power where cigar-chomping businessmen decide our fate. International monetary decisions are made *behind doors*, that are filled with blue **smoke**, where the powerful meet.
Non-anomalous	Few people ever walk down those secret corridors of power where cigar-chomping businessmen decide our fate. International monetary decisions are made *within rooms*, that are filled with blue **smoke**, where the powerful meet.
Question	Are international monetary decisions made by the powerful?

2.	
Anomaly	Item
Anomalous	Many countries have their traditional forms of entertainment. There were specialist female entertainers in *China*, who were traditionally known as **geisha**, and were highly regarded.
Non-anomalous	Many countries have their traditional forms of entertainment. There were specialist female entertainers in *Japan*, who were traditionally known as **geisha**, and were highly regarded.
Question	Do many countries have traditional forms of entertainment?

3.	
Anomaly	Item
Anomalous	The Scottish Tourist Board is trying to attract more tourists to Scotland who want short city breaks. One popular destination is also the *largest* city of Scotland, which is **Glasgow** on the river Clyde.
Non-anomalous	The Scottish Tourist Board is trying to attract more tourists to Scotland who want short city breaks. One popular destination is also the *capital* city of Scotland, which is **Glasgow** on the river Clyde.
Question	Is Glasgow on the river Clyde?

4.	
Anomaly	Item
Anomalous	Historical dramas are very popular on television, especially when they're about the royal family. A popular story is of the *eight* women who ended up marrying **Henry 8th** and is being filmed now.
Non-anomalous	Historical dramas are very popular on television, especially when they're about the royal family. A popular story is of the *six* women who ended up marrying **Henry 8th** and is being filmed now.
Question	Is a historical drama about Henry 8th being made?

5.	
Anomaly	Item
Anomalous	When Dorothy returned from the supermarket she was dismayed to find that her fridge was broken. She checked that she had *plugged in* correctly the appliance's small white **switch** at the back end.
Non-anomalous	When Dorothy returned from the supermarket she was dismayed to find that her fridge was broken. She checked that she had *turned on* correctly the appliance's small white **switch** at the back end.
Question	Was the fridge broken?

6.	
Anomaly	Item
Anomalous	The Inland Revenue are cracking down on criminals thanks to the Chancellor of the Exchequer. They will prosecute anybody who tries to *claim* his new value added **taxes**, and fine them heavily.
Non-anomalous	The Inland Revenue are cracking down on criminals thanks to the Chancellor of the Exchequer. They will prosecute anybody who tries to *evade* his new value added **taxes**, and fine them heavily.
Question	Will people receive heavy fines?

7.	
Anomaly	Item
Anomalous	Meredyth and Derrick had been arguing when she suddenly accused him of having an affair. The shocking accusation just took the *wind* right out of his gaping **mouth** and he fell silent.

Non-anomalous	Meredyth and Derrick had been arguing when she suddenly accused him of having an affair. The shocking accusation just took the *words* right out of his gaping **mouth** and he fell silent.
Question	Had Meredyth and Derek been arguing?

8.	
Anomaly	Item
Anomalous	It has taken Sam Smith three years to release a new album and expectations were high. The majority of music critics couldn't *fault* the new double **album** too highly and it sold well.
Non-anomalous	It has taken Sam Smith three years to release a new album and expectations were high. The majority of music critics couldn't *praise* the new double **album** too highly and it sold well.
Question	Did Sam Smith release an album after 3 years?

9.	
Anomaly	Item
Anomalous	The fairy princess had danced all evening with the handsome prince, but the ball would soon be over. It was so very late when Snow White ran away and left her **slipper** on the stone steps.
Non-anomalous	The fairy princess had danced all evening with the handsome prince, but the ball would soon be over. It was so very late when Cinderella ran away and left her **slipper** on the stone steps.
Question	Had the fairy princess danced all evening?

10.	
Anomaly	Item
Anomalous	The high standards of antenatal and fertility care in NHS hospitals has been praised for its well trained specialists. At the very beginning the pregnant women enter hospital as soon as they **ovulate** and are expertly nursed.
Non-anomalous	The high standards of antenatal and fertility care in NHS hospitals has been praised for its well trained specialists. At the very beginning the hopeful women enter hospital as soon as they **ovulate** and are expertly nursed.
Question	Has care in NHS hospitals been praised?

11.	
Anomaly	Item
Anomalous	Consumer groups are increasingly concerned about highly aggressive sales techniques that can intimidate and anger many innocent people. There are some salesmen who will knock

	aggressively at homeowner's front **doorbells** and are too insistent.
Non-anomalous	Consumer groups are increasingly concerned about highly aggressive sales techniques that can intimidate and anger many innocent people. There are some salesmen who will ring aggressively at homeowner's front **doorbells** and are too insistent.
Question	Are some salesman aggressive?

12.	
Anomaly	Item
Anomalous	Houses are cheaper to heat if they are insulated throughout and a reputable firm has been used. Money is well spent on double glazing of the building's windows and **walls** which helps conserve heat.
Non-anomalous	Houses are cheaper to heat if they are insulated throughout and a reputable firm has been used. Money is well spent on good insulation of the building's windows and **walls** which helps conserve heat.
Question	Are houses cheaper to heat if thoroughly insulated?

13.	
Anomaly	Item
Anomalous	In 1939 the Second World War began and Hitler seemed unstoppable. The turning point came when the Germans attacked the unsuspecting people of **Pearl Harbour** and America declared war.
Non-anomalous	In 1939 the Second World War began and Hitler seemed unstoppable. The turning point came when the Japanese attacked the unsuspecting people of **Pearl Harbour** and America declared war.
Question	Did America declare war?

14.	
Anomaly	Item
Anomalous	When you are ready for a drink, this is how to make the perfect cup of tea. When using a teabag, pour the tea carefully and slowly from the **kettle** into a china cup.
Non-anomalous	When you are ready for a drink, this is how to make the perfect cup of tea. When using a teabag, pour the water carefully and slowly from the **kettle** into a china cup.
Question	Was a china cup used?

15.

Anomaly	Item
Anomalous	The ever-growing popularity of classical music has a fan base around the world that can be seen reflected in mainstream chart success. Many world famous, and very respected singers, have recorded several highly successful **symphonies** that have sold millions.
Non-anomalous	The ever-growing popularity of classical music has a fan base around the world that can be seen reflected in mainstream chart success. Many world famous, and very respected conductors, have recorded several highly successful **symphonies** that have sold millions.
Question	Is classical music growing in popularity?

16.	
Anomaly	Item
Anomalous	It was the biggest ship of its day and no one was prepared for the disaster. On her maiden voyage in the Indian Ocean an accident sunk the **Titanic** in a few hours.
Non-anomalous	It was the biggest ship of its day and no one was prepared for the disaster. On her maiden voyage in the Atlantic Ocean an accident sunk the **Titanic** in a few hours.
Question	Was the titanic on her maiden voyage?

17.	
Anomaly	Item
Anomalous	The long-standing war between the Hutus and Tutsis had been violent, and many people had died. The odds were against the Hutus losing, but even as they faced **defeat**, they raised their weapons.
Non-anomalous	The long-standing war between the Hutus and Tutsis had been violent, and many people had died. The odds were against the Hutus winning, but even as they faced **defeat**, they raised their weapons.
Question	Were weapons raised despite defeat?

18.	
Anomaly	Item
Anomalous	Clive was angry after Amanda called him a fool but didn’t want to lose his temper. During the fight, she had looked at him with such a contemptuous **voice** that he almost cried.
Non-anomalous	Clive was angry after Amanda called him a fool but didn’t want to lose his temper. During the fight, she had spoke to him with such a contemptuous **voice** that he almost cried.
Question	Did Clive almost cry?

19.	
Anomaly	Item
Anomalous	Everybody was disappointed with the University of Leicester team's performance in the end of year relay race. The team were losing, and kept kicking themselves right in the proverbial **foot** by dropping the baton.
Non-anomalous	Everybody was disappointed with the University of Leicester team's performance in the end of year relay race. The team were losing, and kept shooting themselves right in the proverbial **foot** by dropping the baton.
Question	Was everyone not disappointed with University of Leicester's performance?

20.	
Anomaly	Item
Anomalous	Investing money on the stock market can be a high-risk strategy for earning high profits. Money can quickly go up and down like happy children playing on **swings** in the school playground.
Non-anomalous	Investing money on the stock market can be a high-risk strategy for earning high profits. Money can quickly round and round like happy children playing on **swings** in the school playground.
Question	Can investing money in the stock markets be low risk?

21.	
Anomaly	Item
Anomalous	The relationship between American and Russia used to be very frosty. Events changed when visionary leaders of capitalist countries, for example men like **Gorbachev**, opened up their borders.
Non-anomalous	The relationship between American and Russia used to be very frosty. Events changed when visionary leaders of communist countries, for example men like **Gorbachev**, opened up their borders.
Question	Did events stay the same when Gorbachev opened up his borders?

22.	
Anomaly	Item
Anomalous	Recently some multi-faith schools have banned the telling of religious stories that do not reach across different faiths. The story of Jesus on the cross at the time of the **resurrection** has been banned first.

Non-anomalous	Recently some multi-faith schools have banned the telling of religious stories that do not reach across different faiths. The story of Jesus leaving the tomb at the time of the **resurrection** has been banned first.
Question	Have no multi-faith schools banned all religious stories?

23.	
Anomaly	Item
Anomalous	A North American jumbo jet was forced at gunpoint to land in Canada, experts were quickly on hand to help. First of all, the authorities' initial negotiations with the scared and desperate **hostages**, helped calm the situation.
Non-anomalous	A North American jumbo jet was forced at gunpoint to land in Canada, experts were quickly on hand to help. First of all, the authorities' initial communications with the scared and desperate **hostages**, helped calm the situation.
Question	Was a North American jet not forced at gunpoint in Canada?

24.	
Anomaly	Item
Anomalous	A boxing match is meant to last for four rounds, each one lasting for two minutes. Whilst this brutal sport sees many deaths every year, none are actually **serious** and the sport's popular.
Non-anomalous	A boxing match is meant to last for four rounds, each one lasting for two minutes. Whilst this brutal sport sees many injuries every year, none are actually **serious** and the sport's popular.
Question	Is boxing very unpopular?

25.	
Anomaly	Item
Anomalous	The restaurant manager regretfully told Mary that they were very busy and she'd have to wait. He had a table that was currently empty but was likely to be **vacated** in about 20 minutes.
Non-anomalous	The restaurant manager regretfully told Mary that they were very busy and she'd have to wait. He had a table that was currently full but was likely to be **vacated** in about 20 minutes.
Question	Would there be a longer wait than 20 minutes?

26.	
Anomaly	Item

Anomalous	A jockey without a whip at the Grand National is useless. Metaphorically, he is just like a carpenter without his trusty and useful **spanner** vainly attempting the job.
Non-anomalous	A jockey without a whip at the Grand National is useless. Metaphorically, he is just like a plumber without his trusty and useful **spanner** vainly attempting the job.
Question	Is a jockey without a whip very useful?

27.	
Anomaly	Item
Anomalous	When volcanoes erupt the amount of destruction they cause can be terrifying. In one disaster, many people in Rome died because the eruption of **Vesuvius** took them by surprise.
Non-anomalous	When volcanoes erupt the amount of destruction they cause can be terrifying. In one disaster, many people in Pompeii died because the eruption of **Vesuvius** took them by surprise.
Question	Was the disaster expected?

28.	
Anomaly	Item
Anomalous	The Giant Panda bear exists in the wild in only a few isolated places in China. These wild and fragile habitats contain palm trees which **pandas** need for food if they're to survive.
Non-anomalous	The Giant Panda bear exists in the wild in only a few isolated places in China. These wild and fragile habitats contain bamboo trees which **pandas** need for food if they're to survive.
Question	Do giant Pandas not exist in China?

29.	
Anomaly	Item
Anomalous	On Kate Bush's album, Aerial, there is a song about the Catholic saint Joan of Arc. Many think that Joan of Arc was immortal, but she did in fact **exist** in 15th century France.
Non-anomalous	On Kate Bush's album, Aerial, there is a song about the Catholic saint Joan of Arc. Many think that Joan of Arc was fictitious, but she did in fact **exist** in 15th century France.
Question	Do giant Pandas not exist in China?

30.	
Anomaly	Item
Anomalous	It had been a long journey, walking through the forest and carrying a basket of food. Opening the bedroom door the seven dwarves were was surprised to see a **wolf** in the large bed.

Non-anomalous	It had been a long journey, walking through the forest and carrying a basket of food. Opening the bedroom door the red riding hood was surprised to see a **wolf** in the large bed.
Question	Was a wolf not lying in bed?

31.	
Anomaly	Item
Anomalous	The thieves wanted to steal one of the world's most famous paintings from the Parisian museum. Hanging majestically in the Louvre is Rembrandts most famous paintings, the exquisite **Mona Lisa**, and they wanted it.
Non-anomalous	The thieves wanted to steal one of the world's most famous paintings from the Parisian museum. Hanging majestically in the Louvre is Da Vinci most famous paintings, the exquisite **Mona Lisa**, and they wanted it.
Question	Is Mona List not in the Louvre?

32.	
Anomaly	Item
Anomalous	The sociologist was interested in the home life of Britons in the 21st century. He carried out a survey of industrial and, when he could, extended **families**, in five British cities.
Non-anomalous	The sociologist was interested in the home life of Britons in the 21st century. He carried out a survey of nuclear and, when he could, extended **families**, in five British cities.
Question	Was the survey not on British cities?

33.	
Anomaly	Item
Anomalous	Many British holidaymakers want more adventure and are willing to travel further to get it. One of the most popular islands for a long summer holiday is **Brazil** because of the rainforest.
Non-anomalous	Many British holidaymakers want more adventure and are willing to travel further to get it. One of the most popular countries for a long summer holiday is **Brazil** because of the rainforest.
Question	Do holidays makers want less adventure?

34.	
Anomaly	Item
Anomalous	Australia is famous for its flora and fauna. One quintessentially Australian animal is the emu that eats the leaves of **eucalyptus** plants in vast quantities.

Non-anomalous	Australia is famous for its flora and fauna. One quintessentially Australian animal is the koala that eats the leaves of **eucalyptus** plants in vast quantities.
Question	Is Australia not famous for it's flora?

35.	
Anomaly	Item
Anomalous	The naughty schoolboys had been fighting on the muddy playing field and both had black eyes. The headmaster scolded them and they wrung their tired and very muddy **heads** in deeply felt shame.
Non-anomalous	The naughty schoolboys had been fighting on the muddy playing field and both had black eyes. The headmaster scolded them and they hung their tired and very muddy **heads** in deeply felt shame.
Question	Did the headmaster praise the boys?

36.	
Anomaly	Item
Anomalous	Bob and Clare had saved up money all year for a fantastic holiday after they graduated. They travelled all the way to Siberia where the weather was usually **hot** for the whole year.
Non-anomalous	Bob and Clare had saved up money all year for a fantastic holiday after they graduated. They travelled all the way to Ecuador where the weather was usually **hot** for the whole year.
Question	Did Bob and Clare stay at home?

Experimental items for Experiment 3

1.		
Plausibility	Structure	Item
Active	Plausible	It was a warm summer's day. The man walked the hungry and tired dog through the park, in the morning.
Active	Implausible	It was a warm summer's day. The dog walked the hungry and tired| man through the park, in the morning.
Passive	Plausible	It was a warm summer's day. The dog was walked by the hungry and tired| man through the park, in the morning.
Passive	Implausible	It was a warm summer's day. The man was walked by the hungry and tired dog through the park, in the morning.
Question		Who did the pursuing? Policeman | Gangster

1. *		
Plausibility	Structure	Item

Active	Plausible	Crime had increased in the city and people felt unsafe. The policeman pursued the powerful and corrupt gangster but there was no back up.
Active	Implausible	Crime had increased in the city and people felt unsafe. The gangster pursued the powerful and corrupt policeman but there was no back up.
Passive	Plausible	Crime had increased in the city and people felt unsafe. The gangster was pursued by the powerful and corrupt policeman but there as no back up.*
Passive	Implausible	Crime had increased in the city and people felt unsafe. The policeman was pursued by the powerful and corrupt gangster but there as no back up.*
Question		Who did the pursuing? Policeman | Gangster

**This item was deleted due to an error highlighted in red.*

2.		
Plausibility	Structure	Item
Active	Plausible	The battle had lasted for hours. The knight protected the hopeful and loyal residents for as long as possible but the enemy was too strong.
Active	Implausible	The battle had lasted for hours. The residents protected the hopeful and loyal knight for as long as possible but the enemy was too strong.
Passive	Plausible	The battle had lasted for hours. The residents were protected by the hopeful and loyal knight for as long as possible but the enemy was too strong.
Passive	Implausible	The battle had lasted for hours. The knight was protected by the hopeful and loyal residents for as long as possible but the enemy was too strong.
Question		Who did the protecting? Knight | Residents

.

3.		
Plausibility	Structure	
Active	Plausible	The significant trial was widely reported. The lawyer defended the articulate and tenacious victim to the media who were very critical of all who were involved.
Active	Implausible	The significant trial was widely reported. The victim defended the articulate and tenacious lawyer to the media who were very critical of all who were involved.
Passive	Plausible	The significant trial was widely reported. The victim was defended by the articulate and tenacious lawyer to the media who were very critical of all who were involved.
Passive	Implausible	The significant trial was widely reported. The lawyer was defended by the articulate and tenacious victim to the media who were very critical of all who were involved.
Question		Who did the defending? Lawyer | Victim

4.

Plausibility	Structure	
Active	Plausible	The French exam was in a few weeks and more practice was needed. The teacher tested the ambitious and enthusiastic student on his French and found that he did very well.
Active	Implausible	The French exam was in a few weeks and more practice was needed. The student tested the ambitious and enthusiastic teacher on his French and found that he did very well.
Passive	Plausible	The French exam was in a few weeks and more practice was needed. The student was tested by the ambitious and enthusiastic teacher on his French and found that he did very well.
Passive	Implausible	The French exam was in a few weeks and more practice was needed. The teacher was tested by the ambitious and enthusiastic student on his French and found that he did very well.
Question		Who did the testing? Teacher | Student

5.		
Plausibility	Structure	
Active	Plausible	The forest was full of lots of different animals and many people would travel to capture a glimpse of the wildlife. The man hunted the quiet and strong deer for hours.
Active	Implausible	The forest was full of lots of different animals and many people would travel to capture a glimpse of the wildlife. The deer hunted the quiet and strong man for hours.
Passive	Plausible	The forest was full of lots of different animals and many people would travel to capture a glimpse of the wildlife. The deer was hunted by the quiet and strong man for hours.
Passive	Implausible	The forest was full of lots of different animals and many people would travel to capture a glimpse of the wildlife. The man was hunted by the quiet and strong deer for hours.
Question		Who did the hunting? Man | Deer

6.		
Plausibility	Structure	
Active	Plausible	The ballet classes had gone well but there was a competition soon. The instructor guided the thoughtful and experienced ballerina in becoming an award-winning dance professional.
Active	Implausible	The ballet classes had gone well but there was a competition soon. The ballerina guided the thoughtful and experienced instructor in becoming an award-winning dance professional.
Passive	Plausible	The ballet classes had gone well but there was a competition soon. The ballerina was guided by the thoughtful and experienced instructor in becoming an award-winning dance professional.
Passive	Implausible	The ballet classes had gone well but there was a competition soon. The instructor was guided by the thoughtful and

		experienced ballerina in becoming an award-winning dance professional.
Question		Who did the guiding? Ballerina | Instructor

7.		
Plausibility	Structure	
Active	Plausible	Tension was high between the coach and the football player. The coach instructed the clumsy but eager footballer to try a new position that would be better for the team.
Active	Implausible	Tension was high between the coach and the football player. The footballer instructed the clumsy but eager coach to try a new position that would be better for the team.
Passive	Plausible	Tension was high between the coach and the football player. The footballer was instructed by the clumsy but eager coach to try a new position that would be better for the team.
Passive	Implausible	Tension was high between the coach and the football player. The coach was instructed by the clumsy but eager footballer to try a new position that would be better for the team.
Question		Who did the instructing? Coach | Footballer

8.		
Plausibility	Structure	
Active	Plausible	Constant attention was needed on car rides. The mother neglected the dramatic and fussy children momentarily, because of the huge accident that had occurred.
Active	Implausible	Constant attention was needed on car rides. The children neglected the dramatic and fussy mother momentarily, because of the huge accident that had occurred.
Passive	Plausible	Constant attention was needed on car rides. The children were neglected by the dramatic and fussy mother momentarily, because of the huge accident that had occurred.
Passive	Implausible	Constant attention was needed on car rides. The mother was neglected by the dramatic and fussy children momentarily, because of the huge accident that had occurred.
Question		Who did the neglecting? Mother | Children

9.		
Plausibility	Structure	
Active	Plausible	A classical music piece was played in the beautiful De Montfort Hall. The piano played the large and loud musician so well that it brought many to tears.
Active	Implausible	A classical music piece was played in the beautiful De Montfort Hall. The piano was played by the large and loud musician so well that it brought many to tears.
Passive	Plausible	A classical music piece was played in the beautiful De Montfort Hall. The musician was played by the large and loud piano so well that it brought many to tears.

Passive	Implausible	A classical music piece was played in the beautiful De Montfort Hall. The musician played the large and loud piano so well that it brought many to tears.
Question		Who did the playing? Musician | Piano

10.		
Plausibility	Structure	
Active	Plausible	After school, the child often played in the garden. The bird| stroked the young and sweet child tenderly, like a caring mother.
Active	Implausible	After school, the child often played in the garden. The bird was stroked by the young and sweet child tenderly, like a caring mother.
Passive	Plausible	After school, the child often played in the garden. The child was stroked by the young and sweet bird tenderly, like a caring mother.
Passive	Implausible	After school, the child often played in the garden. The child stroked the young and sweet bird tenderly, like a caring mother.
Question		Who did the stroking? Bird | Child

11.		
Plausibility	Structure	
Active	Plausible	The town enjoyed magic shows and often got involved themselves. The audience tricked the naïve and enthusiastic magician with the well-rehearsed and clever tricks.
Active	Implausible	The town enjoyed magic shows and often got involved themselves. The audience was tricked by the naïve and enthusiastic magician with the well-rehearsed and clever tricks.
Passive	Plausible	The town enjoyed magic shows and often got involved themselves. The magician was tricked by the naïve and enthusiastic audience with the well-rehearsed and clever tricks.
Passive	Implausible	The town enjoyed magic shows and often got involved themselves. The magician tricked the naïve and enthusiastic audience with the well-rehearsed and clever tricks.
Question		Who did the tricking? Magician | Audience

12. *		
Plausibility	Structure	
Active	Plausible	The festival lasted days and entertainment between music acts was scheduled. The crowd amused the loud and outlandish comedian with a series of witty jokes.
Active	Implausible	The festival lasted days and entertainment between music acts was scheduled. The crowd was amused by the loud and outlandish comedian with a series of witty jokes.
Passive	Plausible	The festival lasted days and entertainment between music acts was scheduled. The comedian was amused by the loud and outlandish crowd with [a] series of witty jokes.

Passive	Implausible	The festival lasted days and entertainment between music acts was scheduled. The comedian amused the loud and outlandish crowd with a series of witty jokes.
Question		Who did the amusing? Comedian | Crowd

**This item was deleted due to an error highlighted in red.*

13.		
Plausibility	Structure	
Active	Plausible	The pope had many important meetings and some conflicted with each other. The Pope represented the conservative and devoted cardinal at foreign government's engagements.
Active	Implausible	The pope had many important meetings and some conflicted with each other. The Pope was represented by the conservative and devoted cardinal at foreign government's engagements.
Passive	Plausible	The pope had many important meetings and some conflicted with each other. The cardinal was represented by the conservative and devoted Pope at foreign government's engagements.
Passive	Implausible	The pope had many important meetings and some conflicted with each other. The cardinal represented the conservative and devoted Pope at foreign government's engagements.
Question		Who did the representing? Pope | Cardinal

14.		
Plausibility	Structure	
Active	Plausible	The bride had spent hours getting ready. The bride styled the excitable and chatty hairdresser but had to be quick as the ceremony was to start soon.
Active	Implausible	The bride had spent hours getting ready. The bride was styled by the excitable and chatty hairdresser but had to be quick as the ceremony was to start soon.
Passive	Plausible	The bride had spent hours getting ready. The hairdresser was styled by the excitable and chatty bride but had to be quick as the ceremony was to start soon.
Passive	Implausible	The bride had spent hours getting ready. The hairdresser styled the excitable and chatty bride but had to be quick as the ceremony was to start soon.
Question		Who did the styling? Bride | Hairdresser

15.		
Plausibility	Structure	
Active	Plausible	The sheriff and the bandit had once been best friends but it had turned sour. The bandit arrested the outrageous and cruel sheriff after years of searching for him.
Active	Implausible	The sheriff and the bandit had once been best friends but it had turned sour. The bandit was arrested by the outrageous and cruel sheriff after years of searching for him.

Passive	Plausible	The sheriff and the bandit had once been best friends but it had turned sour. The sheriff was arrested by the outrageous and cruel bandit after years of searching for him.
Passive	Implausible	The sheriff and the bandit had once been best friends but it had turned sour. The sheriff arrested the outrageous and cruel bandit after years of searching for him.
Question		Who did the arresting? Bandit | Sheriff

16.		
Plausibility	Structure	
Active	Plausible	Meetings between the counsellor had patient were coming to an end. The patient analysed the difficult and nervous counsellor at every session to consider improvements.
Active	Implausible	Meetings between the counsellor had patient were coming to an end. The patient was analysed by the difficult and nervous counsellor at every session to consider improvements.
Passive	Plausible	Meetings between the counsellor had patient were coming to an end. The counsellor was analysed by the difficult and nervous patient at every session to consider improvements.
Passive	Implausible	Meetings between the counsellor had patient were coming to an end. The counsellor analysed the difficult and nervous patient at every session to consider improvements.
Question		Who did the analysing? Patient | Counsellor

17.		
Plausibility	Structure	
Active	Plausible	The police investigation was nearly resolved, some evidence was still needed. The suspect identified the slim and bald victim almost immediately in the line-up.
Active	Implausible	The police investigation was nearly resolved, some evidence was still needed. The suspect was identified by the slim and bald victim almost immediately in the line-up.
Passive	Plausible	The police investigation was nearly resolved, some evidence was still needed. The victim was identified by the slim and bald suspect almost immediately in the line-up.
Passive	Implausible	The police investigation was nearly resolved, some evidence was still needed. The victim identified the slim and bald suspect almost immediately in the line-up.
Question		Who did the identifying? Suspect | Victim

18. *		
Plausibility	Structure	
Active	Plausible	Heat and smoke made the fireman's job difficult when he went into the building. The person was removed by the brave and strong fireman from the tree just in time.
Active	Implausible	Heat and smoke made the fireman's job difficult when he went into the building. The fireman was removed by the brave and strong person from the tree just in time.

Passive	Plausible	Heat and smoke made the fireman's job difficult when he went into the building. The fireman removed the brave and strong person from the tree just in time.
Passive	Implausible	Heat and smoke made the fireman's job difficult when he went into the called about the building. The person removed the brave and strong fireman from the tree just in time.
Question		Who did the removing? Fireman | Person

**This item was deleted due to an error highlighted in red.*

19. *		
Plausibility	Structure	
Active	Plausible	Every week a panel show was aired where policies were debated. The audience was persuaded by the enthusiastic and passionate panel that the policies would be actioned.
Active	Implausible	Every week a panel show was aired where policies were debated. The panel was persuaded by the enthusiastic and passionate audience that the policies would be actioned.
Passive	Plausible	Every week a panel show was aired where policies were debated. The panel persuaded the enthusiastic and passionate audience that the policies would be actioned.
Passive	Implausible	Every week a panel show was aired where policies were debated. The audience persuaded the enthusiastic and passionate audience panel that the policies would be actioned.
Question		Who did the persuading? Audience | Panel

**This item was deleted due to an error highlighted in red.*

20.		
Plausibility	Structure	
Active	Plausible	A member of the clergy was needed at the Church. The priest was appointed by the faithful and charismatic bishop to perform responsibilities at the church.
Active	Implausible	A member of the clergy was needed at the Church. The bishop was appointed by the faithful and charismatic priest to perform responsibilities at the church.
Passive	Plausible	A member of the clergy was needed at the Church. The bishop appointed the faithful and charismatic priest to perform responsibilities at the church.
Passive	Implausible	A member of the clergy was needed at the Church. The priest appointed the faithful and charismatic bishop to perform responsibilities at the church.
Question		Who did the appointing? Bishop | Priest

21. *		
Plausibility	Structure	
Active	Plausible	The lectures had gotten worse and worse over the term. The student was prohibited by the frank and aggressive student from using a mobile phone during lectures.

Active	Implausible	The lectures had gotten worse and worse over the term. The lecturer was prohibited by the frank and aggressive student from using a mobile phone during lectures.
Passive	Plausible	The lectures had gotten worse and worse over the term. The lecturer prohibited the frank and aggressive student from using a mobile phone during lectures.
Passive	Implausible	The lectures had gotten worse and worse over the term. The student prohibited the frank and aggressive lecturer from using a mobile phone during lectures.
Question		Who did the prohibiting? Lecturer | Student

**This item was deleted due to an error highlighted in red.*

22.		
Plausibility	Structure	
Active	Plausible	The Queen and her staff have a relationship based on mutual respect. The servant was granted by the considerate and brilliant Queen time off for a friend's funeral.
Active	Implausible	The Queen and her staff have a relationship based on mutual respect. The Queen was granted by the considerate and brilliant servant time off for a friend's funeral.
Passive	Plausible	The Queen and her staff have a relationship based on mutual respect. The Queen granted the considerate and brilliant servant time off for a friend's funeral.
Passive	Implausible	The Queen and her staff have a relationship based on mutual respect. The servant granted the considerate and brilliant Queen time off for a friend's funeral.
Question		Who did the granting? Queen | Servant

23.		
Plausibility	Structure	
Active	Plausible	The construction business found apprenticeships to be a huge success. The apprentice was trained by the diligent and driven builder who won employee of the month for his hard work.
Active	Implausible	The construction business found apprenticeships to be a huge success. The builder was trained by the diligent and driven apprentice who won employee of the month for his hard work.
Passive	Plausible	The construction business found apprenticeships to be a huge success. The builder trained the diligent and driven apprentice who won employee of the month for his hard work.
Passive	Implausible	The construction business found apprenticeships to be a huge success. The apprentice trained the diligent and driven builder who won employee of the month for his hard work.
Question		Who did the training? Apprentice | Builder

24.		
Plausibility	Structure	

Active	Plausible	The documentary captured wild animals in their natural habitat. The hyena was chased by the courageous and vicious lion who put up a good fight.
Active	Implausible	The documentary captured wild animals in their natural habitat. The lion was chased by the courageous and vicious hyena who put up a good fight.
Passive	Plausible	The documentary captured wild animals in their natural habitat. The lion chased the courageous and vicious hyena who put up a good fight.
Passive	Implausible	The documentary captured wild animals in their natural habitat. The hyena chased the courageous and vicious lion who put up a good fight.
Question		Who did the chasing? Lion | Hyena

25.		
Plausibility	Structure	
Plausible	Active	The hotel was situated in a convenient place. The traveller was accommodated by the interesting and quirky hotel owner who had trouble lifting the luggage.
Plausible	Passive	The hotel was situated in a convenient place. The hotel owner was accommodated by the interesting and quirky traveller who had trouble lifting the luggage.
Implausible	Active	The hotel was situated in a convenient place. The hotel owner accommodated the interesting and quirky traveller who had trouble lifting the luggage.
Implausible	Passive	The hotel was situated in a convenient place. The traveller accommodated the interesting and quirky hotel owner who had trouble lifting the luggage.
Question		Who did the accommodating? Hotel owner | Traveller

26.		
Plausibility	Structure	
Plausible	Active	Many people waited for weeks to obtain a slot at the beauty parlour. The lady was pampered by the friendly and caring masseuse who was very chatty.
Plausible	Passive	Many people waited for weeks to obtain a slot at the beauty parlour. The masseuse was pampered by the friendly and caring lady who was very chatty.
Implausible	Active	Many people waited for weeks to obtain a slot at the beauty parlour. The masseuse pampered the friendly and caring lady who was very chatty.
Implausible	Passive	Many people waited for weeks to obtain a slot at the beauty parlour. The lady pampered the friendly and caring masseuse who was very chatty.
Question		Who did the pampering? Lady | Masseuse

27.		
Plausibility	Structure	

Plausible	Active	The school had been recognised nationally for record improvements. The governor was gathered by the kind and thoughtful parents in the main hall to relay the good news.
Plausible	Passive	The school had been recognised nationally for record improvements. The governor gathered the kind and thoughtful parents in the main hall to relay the good news.
Implausible	Active	The school had been recognised nationally for record improvements. The parents gathered the kind and thoughtful governors in the main hall to relay the good news.
Implausible	Passive	The school had been recognised nationally for record improvements. The parents were gathered by the kind and thoughtful governors in the main hall to relay the good news.
Question		Who did the gathering? Governor | Parents

28. *		
Plausibility	Structure	
Plausible	Active	The concert was highly anticipated, and it did not disappoint. The fan was praised by the stylish and talented celebrity because the concert was excellent.
Plausible	Passive	The concert was highly anticipated, and it did not disappoint. The fan praised the stylish and talented celebrity because the concert was excellent.
Implausible	Active	The concert was highly anticipated, and it did not disappoint. The celebrity praised the stylish and talented celebrity because the concert was excellent.
Implausible	Passive	The concert was highly anticipated, and it did not disappoint. The celebrity was praised by the stylish and talented fan because the concert was excellent.
Question		Who did the praising? Celebrity | Fan

**This item was deleted due to an error highlighted in red.*

29.		
Plausibility	Structure	
Plausible	Active	The story of the wizard and the King was the children’s favourite. The wizard was bewitched by the respected and kind King with a spell to save his life.
Plausible	Passive	The story of the wizard and the King was the children’s favourite. The wizard bewitched the respected and kind King with a spell to save his life.
Implausible	Active	The story of the wizard and the King was the children’s favourite. The King bewitched the respected and kind wizard with a spell to save his life.
Implausible	Passive	The story of the wizard and the King was the children’s favourite. The King was bewitched by the respected and kind wizard with a spell to save his life.
Question		Who did the bewitching? Wizard | King

30.		
Plausibility	Structure	

Plausible	Active	The newspaper’s sales were low and there was pressure at every level. The editor was criticised by the lazy and boring writer about her work.
Plausible	Passive	The newspaper’s sales were low and there was pressure at every level. The editor criticised the lazy and boring writer about her work.
Implausible	Active	The newspaper’s sales were low and there was pressure at every level. The writer criticised the lazy and boring editor about her work.
Implausible	Passive	The newspaper’s sales were low and there was pressure at every level. The writer was criticised by the lazy and boring editor about her work.
Question		Who did the criticising? Editor | Writer

31.		
Plausibility	Structure	
Plausible	Active	A dragon had been abducting people from Clover Kingdom. The Queen was ordered by the wise and fair princess to save the townspeople.
Plausible	Passive	A dragon had been abducting people from Clover Kingdom. The Queen ordered the serious and wealthy princess to save the townspeople.
Implausible	Active	A dragon had been abducting people from Clover Kingdom. The princess ordered the charming and wealthy Queen to save the townspeople.
Implausible	Passive	A dragon had been abducting people from Clover Kingdom. The princess was ordered by the charming and wealthy Queen to save the townspeople.
Question		Who did the ordering? Queen | Princess

32.		
Plausibility	Structure	
Plausible	Active	The NHS provides free healthcare to people in England, Scotland and Wales. The surgeon was treated by the kind and caring patient back to health, but lots of rest would be needed for a full recovery.
Plausible	Passive	The NHS provides free healthcare to people in England, Scotland and Wales. The surgeon treated the kind and caring patient back to health, but lots of rest would be needed for a full recovery.
Implausible	Active	The NHS provides free healthcare to people in England, Scotland and Wales. The patient treated the kind and caring surgeon back to health, but lots of rest would be needed for a full recovery.
Implausible	Passive	The NHS provides free healthcare to people in England, Scotland and Wales. The patient was treated by the kind and caring surgeon back to health, but lots of rest would be needed for a full recovery.
Question		Who did the treating? Surgeon | Patient

33.		
Plausibility	Structure	
Plausible	Active	The syndicate had authorised many assassinations to keep their illegal business secret. The hitman was killed by the cunning and deceitful target to protect the identities of the secret organisation.
Plausible	Passive	The syndicate had authorised many assassinations to keep their illegal business secret. The hitman killed the cunning and deceitful target to protect the identities of the secret organisation.
Implausible	Active	The syndicate had authorised many assassinations to keep their illegal business secret. The target killed the cunning and deceitful hitman to protect the identities of the secret organisation.
Implausible	Passive	The syndicate had authorised many assassinations to keep their illegal business secret. The target was killed by the cunning and deceitful hitman to protect the identities of the secret organisation.
Question		Who did the killing? Target | Hitman

34.		
Plausibility	Structure	
Plausible	Active	UFO sightings are common across the USA. The alien was abducted by the tall and slender farmer during the night.
Plausible	Passive	UFO sightings are common across the USA. The alien abducted the tall and slender farmer during the night.
Implausible	Active	UFO sightings are common across the USA. The farmer abducted the tall and slender alien during the night.
Implausible	Passive	UFO sightings are common across the USA. The farmer was abducted by the tall and slender alien during the night.
Question		Who did the abducting? Alien | Farmer

35.		
Plausibility	Structure	
Plausible	Active	The police station had one unresolved case. The detective was interrogated by the sincere and sensible witness and managed to solve the case.
Plausible	Passive	The police station had one unresolved case. The detective interrogated the sincere and sensible witness and managed to solve the case.
Implausible	Active	The police station had one unresolved case. The witness interrogated the sincere and sensible detective and managed to solve the case.
Implausible	Passive	The police station had one unresolved case. The witness was interrogated by the sincere and sensible detective and managed to solve the case.
Question		Who did the interrogating? Detective | Witness

Appendix B: Full list of filler stimuli for Experiment 1 and 3.

Filler items for Experiment 1

No.	Item	Question
1	London is a popular place to visit as it is the capital city of England and where the Queen resides. Her Majesty and the royal family responsibly carry out many ceremonial and social duties within the UK and abroad.	Does the Queen reside in England?
2	The world cup is a football tournament that has been played every four years since 1930, except in 1942 and 1946 due to the Second World War. France are the current champions, having won the title in 2018.	Are France current world cup champions?
3	Pollution is defined as the adverse change of a contaminant being introduced into the natural environment. Recycling is one way to reduce pollution, which many individuals in society commit to on a regular basis.	Is recycling one way to reduce pollution?
4	Native American Indians in 1492 lived in a tipi made from buffalo skin, which was an extremely versatile material. In summer the skin could be rolled up to let air in, in winter, adding earth kept the tipi warm.	Is buffalo skin versatile material?
5	In 1947, the partition of India occurred resulting in two self-governing countries, now known as Pakistan and India. This resulted in 14 million refugees. Hostility between these two countries is still present today.	Was the partition of India in 1947?
6	In Greek mythology, Icarus is known for flying too close to the sun and subsequently tumbled into the sea. Icarus's dad made his wings from wax and feathers and warned him about flying too close to the sun or sea.	Did Icarus have wings made of wax and feathers?
7	Festivals have gained much popularity with a range of ages flocking to them. Festivals are a fantastic way to listen to a variety of music, try new foods and enjoy the unique entertainment often put on for everyone to enjoy.	Have festivals become more popular?
8	The profession of a bodyguard is to manage the security and protection of important figures. Recently, one of the prime minister's conference speeches was delayed due to security issues at the location.	Was the prime minister's conference delayed?
9	Often rail strikes in London cause problems for commuters. Passengers have to find alternatives route of travel to reduce the delay to get to work.	Do rail strikes cause problems for commuters?
10	Paris is the capital city of France and is a major European centre for fashion and finance. The couple enjoyed their city break to Paris celebrating their 5th anniversary.	Did the couple enjoy visiting Paris?
11	La Rambla is a vibrant and lively street in Barcelona filled with entertainment, shops and restaurants. A Spanish poet once said "La Rambla was the only street in the world which I wish would never end."	Is La Rambla a vibrant and lively street in Barcelona?

12	Diwali is a Hindu festival that symbolises good over evil or light over dark. The festival of lights is celebrated all over the world with candles, lights and fireworks.	Is the festival of lights celebrated all over the world with candles, lights and fireworks?
13	Chocolate is a popular sweet eaten in many different forms by all ages. The seeds of the cacao are initially very bitter, fermentation allows for the flavour to develop before the seeds are prepared and the shell removed.	Are the seeds of the cacao initially very bitter?
14	Ariya and Adam enjoyed baking cakes of all varieties with their grandma. Their favourite part was once the cake was baked, decorating the cake with icing, sprinkles and fruit.	Did Ariya and Adam enjoy baking cakes?
15	In autumn when there is less sunlight and cooler temperatures, trees shed their leaves. Leaves turn different colours, like orange and red in response to these conditions.	In Autumn, do trees shed their leaves?
16	Poker is a gambling card game that was developed in the 19th century. Poker is played all over the world and is a game that requires strategy and skill.	Is poker played all over the world?
17	In the 18th century, the pump rooms in Royal Leamington spa was a place people went to recuperate because it was believed the water had medicinal powers. Now the Royal pumps hold an art gallery and museum.	Was it believed that the water had medicinal powers?
18	People have enjoyed tattoos for centuries with evidence of tattoos on human mummy remains. Although tattoos are accepted today, less than a 100 years ago, tattoos were associated with criminals, sailors and working men.	Are tattoos accepted today?
19	Tea is the most widely consumed drink in the world after water. This drink is much loved, often offered in homes and hotels as a sign of good hospitality.	Is tea rarely drank?
20	Two presidents in the United States have been impeached, Bill Clinton and Andrew Johnson. The impeachment of Donald Trump is one of great speculation in the UK.	Have more than two presidents been impeached?
21	Ant and Dec are TV presenters who present comedy and entertainment to millions of TV viewers. In recent years, TV is watched live and streamed online.	Are shows no longer streamed online?
22	The competition and markets authority have stated that around 190 billion pound are spent in the UK on groceries each year. Mark and Gail often utilised online shopping to save time.	Do Mark and Gail never use online shopping?
23	The second and third biggest chains in the UK, ASDA and Sainsbury’s plan to merge but an investigation into the merger is underway which is delaying the union.	Are ASDA and Sainsbury's the second and third biggest chains in the UK?
24	The National Health Service is much loved and an important part of British society. Despite the service facing many financial issues, patients are often extremely happy with the service received from NHS staff.	Does the National Health Service have lots of money?

25	The boy was watching his favourite documentary about the planet. He loved learning lots of new facts, for instance, that there are 8 planets due to Pluto no longer being classed as a planet.	Did the boy hate learning about the planets?
26	The woman read the thrilling book during her commute every day on the slow train. Despite the irritating journey she was glad she had something to entertain her.	Did the woman never read her book?
27	The children began every Saturday with swimming lessons. They enjoyed splashing around in the pool, diving and playing games.	Did the children never go swimming?
28	It was only in the 1920's that film was accompanied by sound, prior to this the public would view silent movies. The Jazz singer was the first sound films or "talkies" as they were often referred to.	Is the Jazz Singer a new film?
29	Tim and Becky had a passion for classical music. They would often visit De Montfort Hall to listen to the Philharmonic Orchestra perform their favourite pieces.	Did Tim and Becky hate classical music?
30	The train station was extremely busy that morning due to the strikes and trains were delayed. Tom and Jasmin had decided to visit London to see the famous sites, Buckingham Palace and the London Eye.	Did Tom and Jasmin visit Manchester?
31	Olivia enjoyed singing and had been involved in the choir from a young age. The choir was invited to Norway to sing for Kind Harold and Queen Sonja of Norway, everyone was excited.	Did Olivia dislike singing?
32	To drive in the UK one must hold a drivers licence. Luke had been practising for months and luckily passed first time, with only two minors.	Did Luke pass after the fifth time?
33	One of Hayley and Laura's favourite parts of Christmas was wrapping all the presents. They used lots of ribbon and bows and even made one look like a Christmas cracker.	Did Hayley and Laura dislike wrapping presents?
34	Mr Patel had organised a football club for all senior children at the school. To his surprise, they won their first trophy and everyone was very jubilant with the victory.	Did the children lose?
35	Rosie had a severe nut allergy, therefore always feared travelling by plane. Despite this, she travelled to America and India which were unforgettable holidays.	Did Rosie never fly?
36	The Rogers family have many family members who live all over the world. Thanks to modern technology they regularly video call long distance relatives to keep in contact with them.	Do all the Rogers family members live in one place?

Filler sentences for Experiment 3

No.	Item	Question
1	Every year at Warwick Castle an archery contest is held and Megan and Rob went to watch. The archer relaxed his grip on the bow handle as he concentrated on hitting the target.	What did the archer relax his grip on?
2	People often take a gap year to go exploring. The traveller planned meticulously and strategically his route through the Himalayas.	What did the traveller plan?
3	Most celebrations are celebrated with a cutting of a cake. The baker decorated the birthday cake, carefully and precisely using icing and sprinkles.	What did the baker decorate?
4	Marathons are hard work and lots of training is needed to complete them. The runner sprinted energetically across the finishing line.	What did the runner sprint across?
5	Parks normally consist of lots of vibrant flowers. The gardener planted the wild and colourful flowers in the park for many to enjoy.	What did the gardener plant?
6	Swimming is a common activity on holiday. The boy splashed around in the warm and shallow pool whilst on holiday.	What did the boy splash around in?
7	Now that Keith was retired, he was able to spend more time golfing. The golfer meticulously putted the golf ball resulting in a hole in one.	What did Keith play more of now that he had retired?
8	The guitar is a common instrument that many children learn to play. The boy strummed the expensive guitar so strongly that a string broke!	What did the boy strum?
9	Many different skills are needed to ensure an office works well. The technician fixed the old and slow computers in the office before the office workers began work.	What did the technician fix?
10	Center Parcs is a common destination for families who want to take part in various activities. Kayaking was enjoyed by all the family at Center Parcs, despite all their different interests.	What activity did the family enjoy doing at Center Parcs?
11	The city of Amsterdam is full of rich culture and wealth of history. Amsterdam was explored by bike by Katie and Amy before going to the Anne Frank museum.	What did Katie and Amy explore Amsterdam on?
12	London is the location of the British museum which is dedicated to human history, art and	What did the British Museum exhibit?

	culture. The British museum was chosen to exhibit the Egyptian artefacts for the public to view.	
13	The Lion King musical is the highest grossing Broadway play in history. Going to a musical that is fun and entertaining was James and Karen’s favourite thing to do when they lived in London.	What did Karen and James enjoy seeing in London?
14	The London Eye is the tallest Ferris wheel in Europe. The highest viewing point in London was the London Eye for the public, prior to the construction of The Shard.	What was the highest viewing point in London, prior to the shard?
15	Bournemouth has one of the sandiest beaches in England. Relaxing on the warm sandy beach with the smell of the salty sea was what Kelly enjoyed doing most in summer.	What did Kelly enjoy doing in summer?
16	Bristol is the home of Banksy art and holds a hot air balloon festival every year. The hot air balloons could be seen by thousands of people who watched in Bristol.	What did thousands of people watch?
17	Glasgow cathedral is the only medieval cathedral to survive the protestant reformation in 1560. The tourists were guided by the locals of the best places to see in Glasgow.	What did the tourists see in Glasgow?
18	Every Sunday, Hazel read in bed with her cat Clancy. The Girl with the Dragon Tattoo was what Hazel read on that rainy Sunday morning.	What was read on the rainy Sunday morning?
19	The game was neck and neck. The netball players defended fiercely against the opposition and went on to win the tournament.	What did the netball player win?
20	Films are often adapted from books. The writer typed the amusing screenplay for the movie which everyone hoped would be successful.	What did the writer type?
21	Dance dates back to 3300 BC and is still very common today. The dancer practiced the fluid contemporary piece to impress the judges in the competition.	What did the dancer practice?
22	Devon and Norfolk are two of the best fishing locations in England. The fisherman caught many long and scaly cod for the restaurant owner, which increased his sales.	What did the fisherman catch?
23	Halloween consists of dressing up and going trick-or-treating. The girl told the dark and mysterious story so scarily that all her siblings had nightmares that night.	What did the girl tell?

24	Christmas is the festival that commemorates the birth of Jesus Christ. The parents wrapped the huge toys so well that the children had a hard time unwrapping them on Christmas day!	What did the parents wrap?
25	Remixed songs often include mixes of ~~a~~ new and old melodies. The singer stole the moving song lyrics that songwriter had written 2 years ago without permission which result in a legal action.	What did the songwriter write?
26	Frankie enjoyed shopping for designer clothes. The sales assistant was very helpful to Frankie as she shopped around in Calvin Klein.	
27	Many people seemed to be catching the winter flu. Janice caught the flu and had to leave work due to the terrible fever and congestion she experienced.	What did Janice catch?
28	Every Sunday lunchtime, the family gathered at the grandparents' house. The cupcakes were baked by Ronnie and the grandchildren, thoroughly enjoyed eating them.	What did Ronnie bake?
29	Maisy needed to make alterations to her dress in time for the party. Maisy's dress was sewn by the very talented tailor.	What did the tailor sew?
30	Emma felt terrible about always forgetting to water the plants. All of the roses were secretly watered by Tom who couldn't bear to see his friend kill more plants!	What did Tom water?
31	Amelia and Michelle were best friends and would see each other several times a week. Amelia was visited by Michelle who had been travelling for the past 6 months in Australia.	What had Michelle being doing for the past 6 months?
32	Snowdon is climbed by half a million hikers each year. Snowdon was climbed by the staff members for their local charity, Hope Against Cancer.	What did the staff members climb for charity?
33	Arthur's Seat is 251m above sea level which makes it perfect for seeing excellent views of Edinburgh. Arthur's seat was the location Ben had chosen to propose to Samantha.	What did Ben do in Edinburgh?
34	Cambridge is a beautiful city with lots of history and activities like punting. "Push the pole vertically down!" is what Eleonore was instructed to do by her father.	What did Eleonore do in Cambridge?
35	It was Delilah's first time out to the park with the new puppy. The stick is what the dog tried to ever so precisely and energetically pick up for the owner.	What did the dog try to pick up?

36	Nottingham castle is small but full of lots of history. Robyn visited the castle but it’s closed until 2020 as a 29.4 million pound project is underway.	What is closed?

Appendix C: Experiment 1 – Anomaly detection post hoc analyses

Linear and binomial mixed-effects models post hocs at the pre-critical region.

	Pre-critical region			
Measure	Estimate	SE	t/z value	
Regression path reading times				
Intercept	1416.783	11.533	122.85	***
Older adults: Detected v Non-anomalous	92.377	9.954	9.28	***
Young adults: Detected v Non-anomalous	-39.376	11.601	-3.39	***
Intercept	1411.227	9.269	152.26	***
Older adults: Detected v Undetected	117.896	9.874	11.94	***
Young adults: Detected v Undetected	-167.379	8.224	-20.35	***
Intercept	1422.749	9.657	147.34	***
Older adults: Undetected v Non-anomalous	-60.106	16.33	-3.68	***
Younger adults: Undetected v Non-anomalous	8.458	8.898	0.95	

Regressions in				
Intercept	-0.230	0.100	-2.29	***
Older adults: Detected v Non-anomalous	0.523	0.197	2.66	***
Young adults: Detected v Non-anomalous	0.806	0.231	3.48	***
Intercept	-0.303	0.101	-2.99	***
Older adults: Detected v Undetected	0.063	0.279	0.23	
Young adults: Detected v Undetected	1.186	0.333	3.57	***
Total reading times				
Intercept	1775.989	7.658	231.93	***
Detected: Young v older adults	-146.713	10.058	-14.59	***
Intercept	1806.585	8.261	218.69	***
Older adults: Detected v Non-anomalous	231.618	7.902	29.31	***
Young adults: Detected v Non-anomalous	602.948	7.628	79.05	***
Intercept	1783.876	9.904	180.11	***
Older adults: Detected v Undetected	-131.028	9.444	-13.87	***

Young adults: Detected v Undetected	660.475	10.892	60.64	***

Linear and binomial mixed-effects models post hocs at the critical region.

	Critical region			
Measure	Estimate	SE	t/z value	
First pass reading times				
Intercept	877.524	2.442	359.29	***
Detected anomalies: Young v Older	13.154	2.986	4.41	***
Non-anomalous controls: Young v Older	-17.57	2.153	-8.16	***
Undetected anomalies: Young v Older	-28.486	2.235	-12.75	***
Intercept	875.551	3.172	276.01	***
Older adults: Detected v Non-anomalous	20.721	3.030	6.84	***
Young adults: Detected v Non-anomalous	-45.198	3.480	-12.99	***
Intercept	877.937	2.789	314.81	***

Older adults: Detected v Undetected	57.315	1.997	28.70	***
Young adults: Detected v Undetected	-60.506	3.497	-17.30	***
Regression path reading times				
Intercept	1965.453	3.776	520.54	***
Detected: Young v older adults	-106.171	2.442	-43.47	***
Undetected: Young v older adults	-50.127	4.065	-12.33	***
Non-anomalous: Young v older adults	2.317	2.196	1.06	
Intercept	1969.364	2.339	842.04	***
Older adults: Detected v Undetected	-55.819	2.193	-25.45	***
Young adults: Detected v Undetected	318.172	2.135	149.06	***
Intercept	1956.196	3.005	650.94	***
Older adults: Undetected v Non-anomalous	-153.438	2.939	-52.20	***
Younger adults: Undetected v Non-anomalous	-56.185	2.933	-19.16	***
Regressions in				
Intercept	-0.231	0.045	-5.17	***

Older adults: Detected v Non-anomalous	0.181	0.072	2.51	***
Young adults: Detected v Non-anomalous	0.403	0.086	4.70	***
Total reading times				
Intercept	1872.241	2.533	739.27	***
Older adults: Detected v Non-anomalous	111.859	2.746	40.74	***
Young adults: Detected v Non-anomalous	271.716	3.262	83.29	***
Intercept	1854.018	7.980	232.34	***
Older adults: Detected v Undetected	-39.574	3.015	-13.12	***
Young adults: Detected v Undetected	344.265	3.876	88.81	***
Intercept	1843.152	2.413	763.81	***
Older adults: Non-anomalous v Undetected	64.294	85.909	0.75	
Young adults: Non-anomalous v Undetected	68.641	2.157	31.83	***

Linear and binomial mixed-effects models post hocs at the post-critical region.

	Post-critical region			
Measure	Estimate	SE	t/z value	
First pass reading times				
Intercept	238.707	6.455	36.98	***
Detected: Young v older adults	14.09	6.669	2.113	*
Undetected: Young v older adults	7.561	5.986	1.263	
Non-anomalous: Young v older adults	5.197	7.049	0.737	
Intercept	610.780	14.890	41.02	***
Older adults: Detected v Undetected	-10.650	13.140	-0.81	
Young adults: Detected v Undetected	-148.970	13.400	-11.11	***
Regression path reading times				
Intercept	4539.147	9.109	498.33	***
Detected: Young v older adults	-368.026	11.341	-32.451	***
Undetected: Young v older adults	-122.414	11.096	-11.033	***

Non-anomalous: Young v older adults	148.007	16.082	9.203	***
Intercept	4597.455	8.579	535.88	***
Older adults: Detected v Undetected	-153.229	8.270	-18.53	***
Young adults: Detected v Undetected	1960.585	11.108	176.50	***
Intercept	4447.307	10.988	404.74	***
Older adults: Undetected v Non-anomalous	-492.755	14.159	-34.80	***
Young adults: Undetected v Non-anomalous	-60.098	9.416	1.81	
Total reading times				
Intercept	1405.127	9.822	143.06	***
Older adults: Detected v Non-anomalous	138.712	16.742	8.29	***
Young adults: Detected v Non-anomalous	323.864	11.025	29.38	***
Intercept	1376.824	8.609	159.92	***
Older adults: Detected v Undetected	9.813	8.429	1.16	
Young adults: Detected v Undetected	456.583	8.123	56.21	***

Appendix D – Full list of syntax for final GLM and logistic regression models.

Chapter 4: Experiment 1

Logistic regression syntax for results in Table 7.

At the pre-critical region, the models were: Accuracy ~ First fixation duration, family = binomial, data = Testing. Accuracy ~ First pass reading time, family = binomial, data = Testing. Accuracy ~ Regression path reading time, family = binomial, data = Testing. Accuracy ~ Total reading time, family = binomial, data = Testing. Accuracy ~ Regressions in, family = binomial, data = Testing. Accuracy ~ Regressions out, family = binomial, data = Testing. Accuracy ~ Skipping, family = binomial, data = Testing.

At the critical region, the models were: Accuracy ~ First fixation duration, family = binomial, data = Testing. Accuracy ~ First pass reading time, family = binomial, data = Testing. Accuracy ~ Regression path reading time, family = binomial, data = Testing. Accuracy ~ Total reading time, family = binomial, data = Testing. Accuracy ~ Regressions in, family = binomial, data = Testing. Accuracy ~ Regressions out, family = binomial, data = Testing. Accuracy ~ Skipping, family = binomial, data = Testing.

At the post-critical region, the models were: Accuracy ~ First fixation duration, family = binomial, data = Testing. Accuracy ~ First pass reading time, family = binomial, data = Testing. Accuracy ~ Regression path reading time, family = binomial, data = Testing. Accuracy ~ Total reading time, family = binomial, data = Testing. Accuracy ~ Regressions out, family = binomial, data = Testing.

GLM syntax for results in Table 8.

At the pre-critical region, the GLMM models were: First Fixation time ~ DETECT*AGE + (1+AGE |pp) + (1+ AGE |stim), control=glmercontrol(optimizer= “bobyqa”,optCtrl=list(maxfun=10000)), data = Testing, family= “Gamma”(link = “identity”)). First-pass time ~ DETECT*AGE + (1+AGE |pp) + (1+DETECT|stim), control=glmercontrol(optimizer= “bobyqa”,optCtrl=list(maxfun=10000)), data = Testing, family= “Gamma”(link = “identity”)). Regression path reading time ~ DETECT*AGE + (1 |pp) + (1+ DETECT |stim), control=glmercontrol(optimizer= “bobyqa”,optCtrl=list(maxfun=10000)), data = Testing, family= “Gamma”(link = “identity”)). Total reading time ~ DETECT*AGE + (1+AGE|pp) + (1+AGE|stim), control=glmercontrol(optimizer= “bobyqa”,optCtrl=list(maxfun=10000)), data = Testing, family= “Gamma”(link = “identity”)). Regressions In ~ DETECT*AGE + (1+ AGE |pp) + (1 + DETECT |stim), control=glmercontrol(optimizer= “bobyqa”,optCtrl=list(maxfun=10000)), data = Testing, family= “Gamma”(link = “identity”)).Regressions Out ~ DETECT*AGE + (1|pp) + (1 |stim), control=glmercontrol(optimizer= “bobyqa”,optCtrl=list(maxfun=10000)), data = Testing, family= “Gamma”(link = “identity”)).Probability of skipping ~ DETECT*AGE + (1 |pp) + (1|stim), control=glmercontrol(optimizer= “bobyqa”,optCtrl=list(maxfun=10000)), data = Testing, family= “Gamma”(link = “identity”)).

At the critical region, the GLMM models were: First Fixation time ~ DETECT*AGE + (1+AGE+DETECT |pp) + (1+ AGE+DETECT |stim), control=glmercontrol(optimizer= “bobyqa”,optCtrl=list(maxfun=10000)), data = Testing, family= “Gamma”(link = “identity”)). First-pass reading time ~ DETECT*AGE + (1+AGE |pp) + (1+DETECT|stim), control=glmercontrol(optimizer= “bobyqa”,optCtrl=list(maxfun=10000)), data = Testing, family= “Gamma”(link = “identity”)). Regression path reading time ~ DETECT*AGE + (1+AGE |pp) + (1+DETECT|stim), control=glmercontrol(optimizer= “bobyqa”,optCtrl=list(maxfun=10000)), data = Testing, family= “Gamma”(link = “identity”)). Total reading time ~ DETECT*AGE + (1+AGE|pp) + (1+AGE|stim), control=glmercontrol(optimizer= “bobyqa”,optCtrl=list(maxfun=10000)), data = Testing, family= “Gamma”(link = “identity”)).Regressions In ~ DETECT*AGE + (1+AGE |pp) + (1 +DETECT|stim), control=glmercontrol(optimizer= “bobyqa”,optCtrl=list(maxfun=10000)), data =

Testing, family= “Gamma”(link = “identity”)). Regressions Out ~ DETECT*AGE + (1|pp) + (1+ DETECT |stim), control=glmercontrol(optimizer= “bobyqa”,optCtrl=list(maxfun=10000)), data = Testing, family= “Gamma”(link = “identity”)).Probability of skipping ~ DETECT*AGE + (1+ AGE |pp) + (1+DETECT|stim), control=glmercontrol(optimizer= “bobyqa”,optCtrl=list(maxfun=10000)), data = Testing, family= “Gamma”(link = “identity”)).

At the post-critical region, the GLMM models were: First fixation reading time ~ DETECT*AGE + (1+AGE |pp) + (1+DETECT|stim), control=glmercontrol(optimizer= “bobyqa”,optCtrl=list(maxfun=10000)), data = Testing, family= “Gamma”(link = “identity”)). First pass reading time ~ DETECT*AGE + (1+ DETECT |pp) + (1+AGE+ DETECT |stim), control=glmercontrol(optimizer= “bobyqa”,optCtrl=list(maxfun=10000)), data = Testing, family= “Gamma”(link = “identity”)). Regression path reading time ~ DETECT*AGE + (1+AGE|pp) + (1 |stim), control=glmercontrol(optimizer= “bobyqa”,optCtrl=list(maxfun=10000)), data = Testing, family= “Gamma”(link = “identity”)). Total reading time ~ DETECT *AGE + (1+AGE|pp) + (1+ DETECT |stim), control=glmercontrol(optimizer= “bobyqa”,optCtrl=list(maxfun=10000)), data = Testing, family= “Gamma”(link = “identity”)). Regressions Out ~ DETECT *AGE + (1+AGE|pp) + (1+ AGE |stim), control=glmercontrol(optimizer= “bobyqa”,optCtrl=list(maxfun=10000)), data = Testing, family= “Gamma”(link = “identity”)). Probability of skipping ~ Length*Ambiguity*AGE + (1 |pp) + (1|stim), control=glmercontrol(optimizer= “bobyqa”,optCtrl=list(maxfun=10000)), data = Testing, family= “Gamma”(link = “identity”)).

GLM syntax for results in Table 9.

At the pre-critical region, the GLMM models were: First Fixation time ~ DETECT*AGE + (1+AGE |pp) + (1+ AGE |stim), control=glmercontrol(optimizer= “bobyqa”,optCtrl=list(maxfun=10000)), data = Testing, family= “Gamma”(link = “identity”)). First-pass time ~ DETECT*AGE + (1+AGE |pp) + (1+DETECT|stim), control=glmercontrol(optimizer= “bobyqa”,optCtrl=list(maxfun=10000)), data = Testing, family= “Gamma”(link = “identity”)). Regression path reading time ~ DETECT*AGE + (1 |pp) + (1+ DETECT |stim), control=glmercontrol(optimizer= “bobyqa”,optCtrl=list(maxfun=10000)), data = Testing, family= “Gamma”(link = “identity”)). Total reading time ~ DETECT*AGE + (1+AGE|pp) + (1+DETECT|stim), control=glmercontrol(optimizer= “bobyqa”,optCtrl=list(maxfun=10000)), data = Testing, family= “Gamma”(link = “identity”)). Regressions In ~ DETECT*AGE + (1+ AGE |pp) + (1 |stim), control=glmercontrol(optimizer= “bobyqa”,optCtrl=list(maxfun=10000)), data = Testing, family= “Gamma”(link = “identity”)). Regressions Out ~ DETECT*AGE + (1|pp) + (1+AGE |stim), control=glmercontrol(optimizer= “bobyqa”,optCtrl=list(maxfun=10000)), data = Testing, family= “Gamma”(link = “identity”)).

Probability of skipping ~ DETECT*AGE + (1 |pp) + (1|stim), control=glmercontrol(optimizer= “bobyqa”,optCtrl=list(maxfun=10000)), data = Testing, family= “Gamma”(link = “identity”)).

At the critical region, the GLMM models were: First Fixation time ~ DETECT*AGE + (1+AGE |pp) + (1+ AGE+DETECT |stim), control=glmercontrol(optimizer= “bobyqa”,optCtrl=list(maxfun=10000)), data = Testing, family= “Gamma”(link = “identity”)). First-pass reading time ~ DETECT*AGE + (1+AGE |pp) + (1+DETECT|stim), control=glmercontrol(optimizer= “bobyqa”,optCtrl=list(maxfun=10000)), data = Testing, family= “Gamma”(link = “identity”)). Regression path reading time ~ DETECT*AGE + (1+AGE |pp) + (1+DETECT|stim), control=glmercontrol(optimizer= “bobyqa”,optCtrl=list(maxfun=10000)), data = Testing, family= “Gamma”(link = “identity”)). Total reading time ~ DETECT*AGE + (1+AGE|pp) + (1+DETECT|stim), control=glmercontrol(optimizer= “bobyqa”,optCtrl=list(maxfun=10000)), data = Testing, family= “Gamma”(link = “identity”)). Regressions In ~ DETECT*AGE + (1+AGE |pp) + (1 |stim), control=glmercontrol(optimizer= “bobyqa”,optCtrl=list(maxfun=10000)), data

= Testing, family= “Gamma”(link = “identity”)). Regressions Out ~ DETECT*AGE + (1|pp) + (1|stim), control=glmercontrol(optimizer= “bobyqa”,optCtrl=list(maxfun=10000)), data = Testing, family= “Gamma”(link = “identity”)). Probability of skipping ~ DETECT*AGE + (1+ AGE |pp) + (1+DETECT|stim), control=glmercontrol(optimizer= “bobyqa”,optCtrl=list(maxfun=10000)), data = Testing, family= “Gamma”(link = “identity”)).

At the post-critical region, the GLMM models were: First Fixation time ~ DETECT*AGE + (1+AGE |pp) + (1+ DETECT |stim), control=glmercontrol(optimizer= “bobyqa”,optCtrl=list(maxfun=10000)), data = Testing, family= “Gamma”(link = “identity”)). First-pass time ~ DETECT*AGE + (1+ DETECT |pp) + (1+ AGE |stim), control=glmercontrol(optimizer= “bobyqa”,optCtrl=list(maxfun=10000)), data = Testing, family= “Gamma”(link = “identity”)). Regression path reading time ~ DETECT*AGE + (1+ AGE |pp) + (1|stim), control=glmercontrol(optimizer= “bobyqa”,optCtrl=list(maxfun=10000)), data = Testing, family= “Gamma”(link = “identity”)). Total reading time ~ DETECT*AGE + (1+AGE|pp) + (1 |stim), control=glmercontrol(optimizer= “bobyqa”,optCtrl=list(maxfun=10000)), data = Testing, family= “Gamma”(link = “identity”)). Regressions Out ~ DETECT*AGE + (1+ AGE |pp) + (1+AGE |stim), control=glmercontrol(optimizer= “bobyqa”,optCtrl=list(maxfun=10000)), data = Testing, family= “Gamma”(link = “identity”)).

Chapter 5: Experiment 2

GLM syntax for results in Table 15.

At the pre-critical region: First-pass time ~ Length*Ambiguity*AGE + (1+Length+Ambiguity|pp) + (1+Length+Ambiguity|stim), control=glmercontrol(optimizer= “bobyqa”,optCtrl=list(maxfun=10000)), data = Testing, family= “Gamma”(link = “identity”)). Regression path reading time ~ Length*Ambiguity*AGE + (1+Ambiguity+Length|pp) + (1+Ambiguity+AGE|stim), control=glmercontrol(optimizer= “bobyqa”,optCtrl=list(maxfun=10000)), data = Testing, family= “Gamma”(link = “identity”)). Total reading time ~ Length*Ambiguity*AGE + (1+Ambiguity+AGE|pp) + (1+Ambiguity+AGE|stim), control=glmercontrol(optimizer= “bobyqa”,optCtrl=list(maxfun=10000)), data = Testing, family= “Gamma”(link = “identity”)). Regressions In ~ Length*Ambiguity*AGE + (1+Ambiguity |pp) + (1 +AGE|stim), control=glmercontrol(optimizer= “bobyqa”,optCtrl=list(maxfun=10000)), data = Testing, family= “Gamma”(link = “identity”)). Regressions Out ~ Length*Ambiguity*AGE + (1+AGE+ Ambiguity |pp) + (1+Length|stim), control=glmercontrol(optimizer= “bobyqa”,optCtrl=list(maxfun=10000)), data = Testing, family= “Gamma”(link = “identity”)). Probability of skipping ~ Length*Ambiguity*AGE + (1+AGE|pp) + (1+ Length |stim), control=glmercontrol(optimizer= “bobyqa”,optCtrl=list(maxfun=10000)), data = Testing, family= “Gamma”(link = “identity”)). At the critical region: First-pass reading time ~ Length*Ambiguity*AGE + (1+AGE+Ambiguity|pp) + (1+AGE|stim), control=glmercontrol(optimizer= “bobyqa”,optCtrl=list(maxfun=10000)), data = Testing, family= “Gamma”(link = “identity”)). Regression path reading time ~ Length*Ambiguity*AGE + (1+Ambiguity+Length|pp) + (1+Ambiguity+AGE|stim), control=glmercontrol(optimizer= “bobyqa”,optCtrl=list(maxfun=10000)), data = Testing, family= “Gamma”(link = “identity”)). Total reading time ~ Length*Ambiguity*AGE + (1+Ambiguity*AGE|pp) + (1+Ambiguity+Length|stim), control=glmercontrol(optimizer= “bobyqa”,optCtrl=list(maxfun=10000)), data = Testing, family= “Gamma”(link = “identity”)). Regressions In ~ Length*Ambiguity*AGE + (1+Ambiguity |pp) + (1 +AGE|stim), control=glmercontrol(optimizer= “bobyqa”,optCtrl=list(maxfun=10000)), data = Testing, family= “Gamma”(link = “identity”)). Regressions Out ~ Length*Ambiguity*AGE + (1+AGE|pp) + (1+ Length+Ambiguity|stim), control=glmercontrol(optimizer= “bobyqa”,optCtrl=list(maxfun=10000)), data = Testing, family= “Gamma”(link = “identity”)). Probability of skipping ~ Length*Ambiguity*AGE + (1 |pp) + (1|stim), control=glmercontrol(optimizer= “bobyqa”,optCtrl=list(maxfun=10000)), data = Testing, family= “Gamma”(link = “identity”)). At the post-critical region: First-pass reading time ~ Length*Ambiguity*AGE +

(1+Ambiguity+Length|pp) + (1+Ambiguity+AGE|stim), control=glmercontrol(optimizer= "bobyqa",optCtrl=list(maxfun=10000)), data = Testing, family= "Gamma"(link = "identity")). Regression path reading time ~ Length*Ambiguity*AGE + (1+Ambiguity|pp) + (1+Length|stim), control=glmercontrol(optimizer= "bobyqa",optCtrl=list(maxfun=10000)), data = Testing, family= "Gamma"(link = "identity")). Total reading time ~ Length*Ambiguity*AGE + (1+Ambiguity+AGE|pp) + (1+Ambiguity+Length|stim), control=glmercontrol(optimizer= "bobyqa",optCtrl=list(maxfun=10000)), data = Testing, family= "Gamma"(link = "identity")). Regressions In ~ Length*Ambiguity*AGE + (1+AGE |pp) + (1 +Length|stim), control=glmercontrol(optimizer= "bobyqa",optCtrl=list(maxfun=10000)), data = Testing, family= "Gamma"(link = "identity")). Regressions Out ~ Length*Ambiguity*AGE + (1+AGE+ Ambiguity |pp) + (1+ AGE |stim), control=glmercontrol(optimizer= "bobyqa",optCtrl=list(maxfun=10000)), data = Testing, family= "Gamma"(link = "identity")). Probability of skipping ~ Length*Ambiguity*AGE + (1+ Ambiguity |pp) + (1|stim), control=glmercontrol(optimizer= "bobyqa",optCtrl=list(maxfun=10000)), data = Testing, family= "Gamma"(link = "identity")).

Logistic regression syntax for results in Table 16.

At the pre-critical region, the GLM models were: Accuracy ~ First pass reading time, family = binomial, data = Testing. Accuracy ~ Regression path reading time, family = binomial, data = Testing. Accuracy ~ Total reading time, family = binomial, data = Testing. Accuracy ~ Regressions in, family = binomial, data = Testing. Accuracy ~ Regressions out, family = binomial, data = Testing. Accuracy ~ Skipping, family = binomial, data = Testing. At the critical region, the GLM models were: Accuracy ~ First pass reading time, family = binomial, data = Testing. Accuracy ~ Regression path reading time, family = binomial, data = Testing. Accuracy ~ Total reading time, family = binomial, data = Testing. Accuracy ~ Regressions in, family = binomial, data = Testing. Accuracy ~ Regressions out, family = binomial, data = Testing. Accuracy ~ Skipping, family = binomial, data = Testing. At the post-critical region, the GLM models were: Accuracy ~ First pass reading time, family = binomial, data = Testing. Accuracy ~ Regression path reading time, family = binomial, data = Testing. Accuracy ~ Total reading time, family = binomial, data = Testing. Accuracy ~ Regressions out, family = binomial, data = Testing

Chapter 6: Experiment 3

GLM syntax for results in Table 19.

Glm (Accuracy ~ Progressive saccade length ~ Age, data = Testing)

Glm (Accuracy ~ Regressive saccade ~ Age, data = Testing)

Glm (Accuracy ~ Probability of skipping ~ Age, data = Testing)

GLM syntax for results in Table 21.

Glm (Accuracy ~ Structure*Plausibility*AGE, family = binomial, data = Testing)

GLM syntax for results in Table 22.

Glm (Accuracy ~ Progressive saccade length, family = binomial, data = Testing)

Glm (Accuracy ~ Regressive saccade, family = binomial, data = Testing)

Glm (Accuracy ~ Probability of skipping, family = binomial, data = Testing)

GLM syntax for results in Table 23.

Glm (Accuracy ~ Progressive saccade length *Age, family = binomial, data = Testing)

Glm (Accuracy ~ Regressive saccade*Age, family = binomial, data = Testing)

Glm (Accuracy ~ Probability of skipping*Age, family = binomial, data = Testing)

GLM syntax for results in Table 27.

GLMM model: Accuracy ~ Context*Reference*Age + (1+ Context | PP) + (1+Context | item), control=glmercontrol(optimizer= "bobyqa",optCtrl=list(maxfun=10000)), data = Testing, family= "binomial").

Chapter 7: Experiment 4

GLM syntax for results in Table 28.

At the critical region, the GLMM models were:

First Fixation time ~ Context*Reference*Age + (1+ Age | PP) + (1+Context | item), control=glmercontrol(optimizer= "bobyqa",optCtrl=list(maxfun=10000)), data = Testing, family= "Gamma"(link = "identity")).

First-pass time ~ Context*Reference*Age + (1+ Reference| PP) + (1+Reference | item), control=glmercontrol(optimizer= "bobyqa",optCtrl=list(maxfun=10000)), data = Testing, family= "Gamma"(link = "identity")).

Regression path reading time ~ Context*Reference*Age + (1+ Context+Reference| PP) + (1+Reference | item), control=glmercontrol(optimizer= "bobyqa",optCtrl=list(maxfun=10000)), data = Testing, family= "Gamma"(link = "identity")).

Total reading time ~ Context*Reference*Age + (1+ Context+Reference| PP) + (1+Context+Reference | item), control=glmercontrol(optimizer= "bobyqa",optCtrl=list(maxfun=10000)), data = Testing, family= "Gamma"(link = "identity")).

At the post-critical region, the GLMM models were:

First Fixation time ~ Context*Reference*Age + (1+ Context | PP) + (1+Context | item), control=glmercontrol(optimizer= "bobyqa",optCtrl=list(maxfun=10000)), data = Testing, family= "Gamma"(link = "identity")).

First-pass reading time ~ Context*Reference*Age + (1+ Context+Age | PP) + (1+Context | item), control=glmercontrol(optimizer= "bobyqa",optCtrl=list(maxfun=10000)), data = Testing, family= "Gamma"(link = "identity")).

Regression path reading time ~ Context*Reference*Age + (1+ Reference | PP) + (1+ Reference | item), control=glmercontrol(optimizer= "bobyqa",optCtrl=list(maxfun=10000)), data = Testing, family= "Gamma"(link = "identity")).

Total reading time ~ Context*Reference*Age + (1+ Context+Reference | PP) + (1+Context+Reference | item), control=glmercontrol(optimizer= "bobyqa",optCtrl=list(maxfun=10000)), data = Testing, family= "Gamma"(link = "identity")).

GLM syntax for results in Table 29.

Frequency covariate included in model:

First Fixation time ~ Context*Reference*Age + Frequency + (1+Context + Reference + Age | PP) + (1+Context + Reference + Age | item), control=glmercontrol(optimizer= “bobyqa”,optCtrl=list(maxfun=10000)), data = Testing, family= “Gamma”(link = “identity”)).

This model was the same for first-pass reading times.

Length covariate included in model:

First Fixation time ~ Context*Reference*Age + Length + (1+Context + Reference | PP) + (1+Context + Reference + AGE | item),

control=glmercontrol(optimizer= “bobyqa”,optCtrl=list(maxfun=10000)), data = Testing, family= “Gamma”(link = “identity”))

First-pass reading time ~ Context*Reference*Age + Length + (1+Context | PP) + (1+Context + Reference | item),

control=glmercontrol(optimizer= “bobyqa”,optCtrl=list(maxfun=10000)), data = Testing, family= “Gamma”(link = “identity”))

Appendix E: Full list of experimental stimuli for Experiment 2 and 4.

Experimental trials for Experiment 2

1.		
Length	Ambiguity	Item
Long	Ambiguous	The mother entered the room. While the child was painting a truck that was quite large was slowly being loaded in a garage nearby. The mother told her son that dinner was ready.
Long	Unambiguous	The mother entered the room. While the child was painting, a truck that was quite large was slowly being loaded in a garage nearby. The mother told her son that dinner was ready.
Short	Ambiguous	The mother entered the room. While the child was painting a truck was slowly being loaded in a garage nearby. The mother told her son that dinner was ready.
Short	Unambiguous	The mother entered the room. While the child was painting, a truck was slowly being loaded in a garage nearby. The mother told her son that dinner was ready.
Question		Was the child painting a truck? Y | **N**

2.		
Length	Ambiguity	Item
Long	Ambiguous	Many documents in the office were secret. As the chairman dictated the report that was very worrying was quickly being destroyed in the other room. The office was to be inspected on Monday.
Long	Unambiguous	Many documents in the office were secret. As the chairman dictated, the report that was very worrying was quickly being destroyed in the other room. The office was to be inspected on Monday.
Short	Ambiguous	Many documents in the office were secret. As the chairman dictated the report was quickly being destroyed in the other room. The office was to be inspected on Monday.
Short	Unambiguous	Many documents in the office were secret. As the chairman dictated, the report was quickly being destroyed in the other room. The office was to be inspected on Monday.
Question		Was the chairman dictating the report? Y | **N**

3.		
Length	Ambiguity	Item

Long	Ambiguous	The air in the classroom was very bad. While the old teacher was smoking the cigarettes that were smelly were hastily being extinguished by the pupils. The next lesson was Mathematics.
Long	Unambiguous	The air in the classroom was very bad. While the old teacher was smoking, the cigarettes that were smelly were hastily being extinguished by the pupils. The next lesson was Mathematics.
Short	Ambiguous	The air in the classroom was very bad. While the old teacher was smoking the cigarettes were hastily being extinguished by the pupils. The next lesson was Mathematics.
Short	Unambiguous	The air in the classroom was very bad. While the old teacher was smoking, the cigarettes were hastily being extinguished by the pupils. The next lesson was Mathematics.
Question		Was the old teacher smoking the cigarettes? Y | **N**

4.		
Length	Ambiguity	Item
Long	Ambiguous	Chaos broke out on the battlefield. As the warrior was fighting the tyrant who was badly injured was slowly being carried away by his slaves. The fighting would be at an end soon.
Long	Unambiguous	Chaos broke out on the battlefield. As the warrior was fighting, the tyrant who was badly injured was slowly being carried away by his slaves. The fighting would be at an end soon.
Short	Ambiguous	Chaos broke out on the battlefield. As the warrior was fighting the tyrant was slowly being carried away by his slaves. The fighting would be at an end soon.
Short	Unambiguous	Chaos broke out on the battlefield. As the warrior was fighting, the tyrant was slowly being carried away by his slaves. The fighting would be at an end soon.
Question		Was the warrior fighting the tyrant? Y | **N**

5.		
Length	Ambiguity	Item
Long	Ambiguous	The waitress had lots of work to do in the restaurant. While the gentleman was eating the burgers that were really huge were still being reheated in the microwave. The gentleman contemplated whether to order dessert.
Long	Unambiguous	The waitress had lots of work to do in the restaurant. While the gentleman was eating, the burgers that were really huge were still being reheated in the microwave. The gentleman contemplated whether to order dessert.
Short	Ambiguous	The waitress had lots of work to do in the restaurant. While the gentleman was eating the burgers were still being reheated in the microwave. The gentleman contemplated whether to order dessert.
Short	Unambiguous	The waitress had lots of work to do in the restaurant. While the gentleman was eating, the burgers were still being reheated in

		the microwave. The gentleman contemplated whether to order dessert.
Question		Was the gentleman eating the burgers?

6.		
Length	Ambiguity	Item
Long	Ambiguous	It was a cold morning on the field. While the team captain trained the goalkeeper who was very agile was already being bandaged in the dressing room. Football fans would shortly arrive.
Long	Unambiguous	It was a cold morning on the field. While the team captain trained, the goalkeeper who was very agile was already being bandaged in the dressing room. Football fans would shortly arrive.
Short	Ambiguous	It was a cold morning on the field. While the team captain trained the goalkeeper was already being bandaged in the dressing room. Football fans would shortly arrive.
Short	Unambiguous	It was a cold morning on the field. While the team captain trained, the goalkeeper was already being bandaged in the dressing room. Football fans would shortly arrive.
Question		Was the team captain training the goalkeeper?

7.		
Length	Ambiguity	Item
Long	Ambiguous	It was a beautiful evening in the harbour. As the skipper was sailing the ship that was quite small was still being repaired at the dock. The fish market was due a delivery on Thursday.
Long	Unambiguous	It was a beautiful day in the harbour. As the skipper was sailing, the ship that was quite small was still being repaired at the dock. The fish market was due a delivery on Thursday.
Short	Ambiguous	It was a beautiful day in the harbour. As the skipper was sailing the ship was still being repaired at the dock. The fish market was due a delivery on Thursday.
Short	Unambiguous	It was a beautiful day in the harbour. As the skipper was sailing, the ship was still being repaired at the dock. The fish market was due a delivery on Thursday.
Question		Was the skipper sailing the ship?

8.		
Length	Ambiguity	Item
Long	Ambiguous	Most of the class got up too late. While the professor lectured the students who were very young were still being driven to the lecture hall. This was not a good start to the term.
Long	Unambiguous	Most of the class got up too late. While the professor lectured, the students who were very young were still being driven to the lecture hall. This was not a good start to the term.

Short	Ambiguous	Most of the class got up too late. While the professor lectured the students were still being driven to the lecture hall. This was not a good start to the term.
Short	Unambiguous	Most of the class got up too late. While the professor lectured, the students were still being driven to the lecture hall. This was not a good start to the term.
Question		Was the professor lecturing the students?

9.		
Length	Ambiguity	Item
Long	Ambiguous	It was very busy in the restaurant. While the cook was chopping the tomatoes that were very red were carefully being sorted by the waiter. The ingredients would be perfect for the pasta dish.
Long	Unambiguous	It was very busy in the restaurant. While the cook was chopping, the tomatoes that were very red were carefully being sorted by the waiter. The ingredients would be perfect for the pasta dish.
Short	Ambiguous	It was very busy in the restaurant. While the cook was chopping the tomatoes were carefully being sorted by the waiter. The ingredients would be perfect for the pasta dish.
Short	Unambiguous	It was very busy in the restaurant. While the cook was chopping, the tomatoes were carefully being sorted by the waiter. The ingredients would be perfect for the pasta dish.
Question		Was the cook chopping the tomatoes?

10.		
Length	Ambiguity	Item
Long	Ambiguous	It was a Tuesday in the office. As the salesman was phoning the assistants who were quite hard-working were severely being criticised in the other room. This would affect staff morale.
Long	Unambiguous	It was a Tuesday in the office. As the salesman was phoning, the assistants who were quite hard-working were severely being criticised in the other room. This would affect staff morale.
Short	Ambiguous	It was a Tuesday in the office. As the salesman was phoning the assistants were severely being criticised in the other room. This would affect staff morale.
Short	Unambiguous	It was a Tuesday in the office. As the salesman was phoning, the assistants were severely being criticised in the other room. This would affect staff morale.
Question		Was the salesman phoning the assistants?

11.		
Length	Ambiguity	Item

Long	Ambiguous	The ranger noticed aggressive animals near the desert camp. As the lion attacked the elephant that was young was already being transported to the zoo. The truck's tyres were covered in sand.
Long	Unambiguous	The ranger noticed aggressive animals near the desert camp. As the lion attacked, the elephant that was young was already being transported to the zoo. The truck's tyres were covered in sand.
Short	Ambiguous	The ranger noticed aggressive animals near the desert camp. As the lion attacked the elephant was already being transported to the zoo. The truck's tyres were covered in sand.
Short	Unambiguous	The ranger noticed aggressive animals near the desert camp. As the lion attacked, the elephant was already being transported to the zoo. The truck's tyres were covered in sand.
Question		Did the lion attack the elephant?

12.		
Length	Ambiguity	Item
Long	Ambiguous	The bake sale was always well attended. While the woman was baking the cookies that were very sweet were safely being stored in the fridge. The group wanted to raise lots of money for charity.
Long	Unambiguous	The bake sale was always well attended. While the woman was baking, the cookies that were very sweet were safely being stored in the fridge. The group wanted to raise lots of money for charity.
Short	Ambiguous	The bake sale was always well attended. While the woman was baking the cookies were safely being stored in the fridge. The group wanted to raise lots of money for charity.
Short	Unambiguous	The bake sale was always well attended. While the woman was baking, the cookies were safely being stored in the fridge. The group wanted to raise lots of money for charity.
Question		Was the woman baking the cookies?

13.		
Length	Ambiguity	Item
Long	Ambiguous	There were lots of vehicles at the filling station. While the driver was parking the motorcycle that was expensive was already being repaired by the mechanic. There would be a long wait for petrol.
Long	Unambiguous	There were lots of vehicles at the filling station. While the driver was parking, the motorcycle that was expensive was already being repaired by the mechanic. There would be a long wait for petrol.
Short	Ambiguous	There were lots of vehicles at the filling station. While the driver was parking the motorcycle was already being repaired by the mechanic. There would be a long wait for petrol.

Short	Unambiguous	There were lots of vehicles at the filling station. While the driver was parking, the motorcycle was already being repaired by the mechanic. There would be a long wait for petrol.
Question		Was the driver parking the motorcycle?

14.		
Length	Ambiguity	Item
Long	Ambiguous	The nursery was preparing a celebration. While the young girl was drawing flowers which were very beautiful were carefully being arranged for the party. There was a lot to do before the guests arrived.
Long	Unambiguous	The nursery was preparing a celebration. While the young girl was drawing, flowers which were very beautiful were carefully being arranged for the party. There was a lot to do before the guests arrived.
Short	Ambiguous	The nursery was preparing a celebration. While the young girl was drawing flowers were carefully being arranged for the party. There was a lot to do before the guests arrived.
Short	Unambiguous	The nursery was preparing a celebration. While the young girl was drawing, flowers were carefully being arranged for the party. There was a lot to do before the guests arrived.
Question		Was the young girl drawing the flowers?

15.		
Length	Ambiguity	Item
Long	Ambiguous	It was a very hectic afternoon in the call centre. When a customer called the manager who was organised was harshly being told to be quiet by the agent on the phone. New staff had been employed recently.
Long	Unambiguous	It was a very hectic afternoon in the call centre. When a customer called, the manager who was organised was harshly being told to be quiet by the agent on the phone. New staff had been employed recently.
Short	Ambiguous	It was a very hectic afternoon in the call centre. When a customer called the manager was harshly being told to be quiet by the agent on the phone. New staff had been employed recently.
Short	Unambiguous	It was a very hectic afternoon in the call centre. When a customer called, the manager was harshly being told to be quiet by the agent on the phone. New staff had been employed recently.
Question		Did the customer call the manager?

16.		
Length	Ambiguity	Item
Long	Ambiguous	Many people were hungry at the beach. While the boy was eating the ice cream which was quite tasty was quickly being sold to the tourists. The beach was covered in litter.

Long	Unambiguous	Many people were hungry at the beach. While the boy was eating, the ice cream which was quite tasty was quickly being sold to the tourists. The beach was covered in litter.
Short	Ambiguous	Many people were hungry at the beach. While the boy was eating the ice cream which was quite tasty was quickly being sold to the tourists. The beach was covered in litter.
Short	Unambiguous	Many people were hungry at the beach. While the boy was eating, the ice cream was quickly being sold to the tourists. The beach was covered in litter.
Question		Was the boy eating the ice cream?

17.		
Length	Ambiguity	Item
Long	Ambiguous	Many people get thirsty at night. While the girl was drinking the milk which was very cold was mysteriously being removed from the fridge. Dishes from the evening lay in the sink.
Long	Unambiguous	Many people get thirsty at night. While the girl was drinking, the milk which was very cold was mysteriously being removed from the fridge. Dishes from the night before lay in the sink.
Short	Ambiguous	Many people get thirsty at night. While the girl was drinking the milk was mysteriously being removed from the fridge. Dishes from the night before lay in the sink.
Short	Unambiguous	Many people get thirsty at night. While the girl was drinking, the milk was mysteriously being removed from the fridge. Dishes from the night before lay in the sink.
Question		Was the girl drinking the milk?

18.		
Length	Ambiguity	Item
Long	Ambiguous	The firm dealt with lots of complicated business deals. While the lawyer was studying the contracts which were important were safely being stored in the cupboard. This case would take some time to settle.
Long	Unambiguous	The firm dealt with lots of complicated business deals. While the lawyer was studying, the contracts which were important were safely being stored in the cupboard. This case would take some time to settle.
Short	Ambiguous	The firm dealt with lots of complicated business deals. While the lawyer was studying the contracts were safely being stored in the cupboard. This case would take some time to settle.
Short	Unambiguous	The firm dealt with lots of complicated business deals. While the lawyer was studying, the contracts were safely being stored in the cupboard. This case would take some time to settle.
Question		Was the lawyer studying the contracts?

19.		
Length	Ambiguity	Item

Long	Ambiguous	It was the first day of the wrestling tournament. As the newcomer wrestled the athlete who was very good was carefully being treated by the medics. The newcomer would go on to the second day of the tournament.
Long	Unambiguous	It was the first day of the wrestling tournament. As the newcomer wrestled, the athlete who was very good was carefully being treated by the medics. The newcomer would go on to the second day of the tournament.
Short	Ambiguous	It was the first day of the wrestling tournament. As the newcomer wrestled the athlete was carefully being treated by the medics. The newcomer would go on to the second day of the tournament.
Short	Unambiguous	It was the first day of the wrestling tournament. As the newcomer wrestled, the athlete was carefully being treated by the medics. The newcomer would go on to the second day of the tournament.
Question		Did the newcomer wrestle the athlete?

20.		
Length	Ambiguity	Item
Long	Ambiguous	The forest was not as quiet as it seemed. While the painter was sketching the trees that were quite tall were quickly being burnt down elsewhere in the wood. Firefighters were on their way.
Long	Unambiguous	The forest was not as quiet as it seemed. While the painter was sketching, the trees that were quite tall were quickly being burnt down elsewhere in the wood. Firefighters were on their way.
Short	Ambiguous	The forest was not as quiet as it seemed. While the painter was sketching the trees were quickly being burnt down elsewhere in the wood. Firefighters were on their way.
Short	Unambiguous	The forest was not as quiet as it seemed. While the painter was sketching, the trees were quickly being burnt down elsewhere in the wood. Firefighters were on their way.
Question		Was the painter sketching the trees?

21.		
Length	Ambiguity	Item
Long	Ambiguous	The artist had not drawn in some years. As the artist drew the portrait that was very expensive was being stolen by thieves. Thieves target many art exhibitions.
Long	Unambiguous	The artist had not drawn in some years. As the artist drew, the portrait that was very expensive was being stolen by thieves. Thieves target many art exhibitions.
Short	Ambiguous	The artist had not drawn in some years. As the artist drew the portrait was being stolen by thieves. Thieves target many art exhibitions.

Short	Unambiguous	The artist had not drawn in some years. As the artist drew, the portrait was being stolen by thieves. Thieves target many art exhibitions.
Question		Was the artist drawing the portrait?

22.		
Length	Ambiguity	Item
Long	Ambiguous	The guitarist had been practicing for the show. While the musician was playing the guitar that was a bit old was carefully being retuned in the studio. The music had to be perfect for the concert.
Long	Unambiguous	The guitarist had been practicing for the show. While the musician was playing, the guitar that was a bit old was carefully being retuned in the studio. The music had to be perfect for the concert.
Short	Ambiguous	The guitarist had been practicing for the show. While the musician was playing the guitar was carefully being retuned in the studio. The music had to be perfect for the concert.
Short	Unambiguous	The guitarist had been practicing for the show. While the musician was playing, the guitar was carefully being retuned in the studio. The music had to be perfect for the concert.
Question		Was the musician playing the guitar?

23.		
Length	Ambiguity	Item
Long	Ambiguous	There were many deadlines for the students on the course to meet. As the student typed the essay that was very long was being handed in for submission. Some students like to celebrate after submitting an assignment.
Long	Unambiguous	There were many deadlines for the students on the course to meet. As the student typed, the essay that was very long was being handed in for submission. Some students like to celebrate after submitting an assignment.
Short	Ambiguous	There were many deadlines for the students on the course to meet. As the student typed the essay was being handed in for submission. Some students like to celebrate after submitting an assignment.
Short	Unambiguous	There were many deadlines for the students on the course to meet. As the student typed, the essay was being handed in for submission. Some students like to celebrate after submitting an assignment.
Question		Was the student typing the essay?

24.		
Length	Ambiguity	Item
Long	Ambiguous	The talent show was always well attended. While the magician was practicing the routine that was very good was comically

		being performed in the hall. A magician had never won the talent show before.
Long	Unambiguous	The talent show was always well attended. While the magician was practicing, the routine that was very good was comically being performed in the hall. A magician had never won the talent show before.
Short	Ambiguous	The talent show was always well attended. While the magician was practicing the routine was comically being performed in the hall. A magician had never won the talent show before.
Short	Unambiguous	The talent show was always well attended. While the magician was practicing, the routine was comically being performed in the hall. A magician had never won the talent show before.
Question		Was the magician practicing the routine?

25.		
Length	Ambiguity	Item
Long	Ambiguous	The school was highly subscribed. While the children learned the play that was very funny was being practiced in the school hall. Many activities were organised for all ages.
Long	Unambiguous	The school was highly subscribed. While the children learned, the play that was very funny was being practiced in the school hall. Many activities were organised for all ages.
Short	Ambiguous	The school was highly subscribed. While the children learned the play was being practiced in the school hall. Many activities were organised for all ages.
Short	Unambiguous	The school was highly subscribed. While the children learned, the play was being practiced in the school hall. Many activities were organised for all ages.
Question		Were the children learning the play?

26.		
Length	Ambiguity	Item
Long	Ambiguous	There were many attractions in the city. As the tourist explored the castle that was very spooky was slowly being renovated until the end of the year. The castle was very old.
Long	Unambiguous	There were many attractions in the city. As the tourist explored, the castle that was very spooky was slowly being renovated until the end of the year. The castle was very old.
Short	Ambiguous	There were many attractions in the city. As the tourist explored the castle was slowly being renovated until the end of the year. The castle was very old.
Short	Unambiguous	There were many attractions in the city. As the tourist explored, the castle was slowly being renovated until the end of the year. The castle was very old.
Question		Did the tourist explore the castle?

27.		
Length	Ambiguity	Item

Long	Ambiguous	The student liked revising in the library. As the student studied the book that was quite thick was clumsily being dropped on the floor. It was a popular place during exam periods.
Long	Unambiguous	The student liked revising in the library. As the student studied, the book that was quite thick was clumsily being dropped on the floor. It was a popular place during exam periods.
Short	Ambiguous	The student liked revising in the library. As the student studied the book was clumsily being dropped on the floor. It was a popular place during exam periods.
Short	Unambiguous	The student liked revising in the library. As the student studied, the book was clumsily being dropped on the floor. It was a popular place during exam periods.
Question		Did the student study the book?

28.		
Length	Ambiguity	Item
Long	Ambiguous	The man enjoyed reading every day. As the man read the news that was quite shocking was rapidly being conveyed to all who needed to know. The news would take some time to sink in.
Long	Unambiguous	The man enjoyed reading every day. As the man read, the news that was quite shocking was rapidly being conveyed to all who needed to know. The news would take some time to sink in.
Short	Ambiguous	The man enjoyed reading every day. As the man read the news was rapidly being conveyed to all who needed to know. The news would take some time to sink in.
Short	Unambiguous	The man enjoyed reading every day. As the man read, the news was rapidly being conveyed to all who needed to know. The news would take some time to sink in.
Question		Did the man read the news?

29.		
Length	Ambiguity	Item
Long	Ambiguous	Everyone was excited for Christmas. While the kids played the board game that was quite fun was secretly being wrapped by their parents. There would be lots of presents under the tree.
Long	Unambiguous	Everyone was excited for Christmas. While the kids played, the board game that was quite fun was secretly being wrapped by their parents. There would be lots of presents under the tree.
Short	Ambiguous	Everyone was excited for Christmas. While the kids played the board game was secretly being wrapped by their parents. There would be lots of presents under the tree.
Short	Unambiguous	Everyone was excited for Christmas. While the kids played, the board game was secretly being wrapped by their parents. There would be lots of presents under the tree.
Question		Were the kids playing with the board game?

30.		
Length	Ambiguity	Item

Long	Ambiguous	The wedding outfits needed final adjustments. While the tailor sewed the zip that was very tricky was cautiously being slid up to the top of the dress. Mary would need to be careful when dancing.
Long	Unambiguous	The wedding outfits needed final adjustments. While the tailor sewed, the zip that was very tricky was cautiously being slid up to the top of the dress. Mary would need to be careful when dancing.
Short	Ambiguous	The wedding outfits needed final adjustments. While the tailor sewed the zip was cautiously being slid up to the top of the dress. Mary would need to be careful when dancing.
Short	Unambiguous	The wedding outfits needed final adjustments. While the tailor sewed, the zip was cautiously being slid up to the top of the dress. Mary would need to be careful when dancing.
Question		Was the tailor sewing the zip?

31.		
Length	Ambiguity	Item
Long	Ambiguous	Everyone had arrived for the celebration on time. As the barman served the champagne that was very bubbly was carefully being poured into flutes for the toast. Diane’s party was going well.
Long	Unambiguous	Everyone had arrived for the celebration on time. As the barman served, the champagne that was very bubbly was carefully being poured into flutes for the toast. Diane’s party was going well.
Short	Ambiguous	Everyone had arrived for the celebration on time. As the barman served the champagne was carefully being poured into flutes for the toast. Diane’s party was going well.
Short	Unambiguous	Everyone had arrived for the celebration on time. As the barman served, the champagne was carefully being poured into flutes for the toast. Diane’s party was going well.
Question		Was the barman serving the champagne?

32.		
Length	Ambiguity	Item
Long	Ambiguous	There was a build up of traffic. As the policeman was signalling the people who were quite frustrated were quickly walking into the shops. It was a very busy day in the town centre.
Long	Unambiguous	There was a build up of traffic. As the policeman was signalling, the people who were quite frustrated were quickly walking into the shops. It was a very busy day in the town centre.
Short	Ambiguous	There was a build up of traffic. As the policeman was signalling the people were quickly walking into the shops. It was a very busy day in the town centre.

Short	Unambiguous	There was a build up of traffic. As the policeman was signalling, the people were quickly walking into the shops. It was a very busy day in the town centre.
Question		Was the policeman signalling the people?

Experimental items for Experiment 4

1.		
Context	Reference	Item
Singular	Singular	Every room at the local bed and breakfast has a bathroom. This bathroom regularly gets inspected by the landlady. The bed and breakfast is a reputable and popular establishment.
Singular	Plural	Every room at the local bed and breakfast has a bathroom. These bathrooms regularly get inspected by the landlady. The bed and breakfast is a reputable and popular establishment.
Plural	Singular	Every visitor at the local bed and breakfast signs a guestbook. This guestbook regularly gets inspected by the landlady. The bed and breakfast is a reputable and popular establishment.
Plural	Plural	Every visitor at the local bed and breakfast signs a guestbook. These guestbooks regularly get inspected by the landlady. The bed and breakfast is a reputable and popular establishment.
Question		Does the landlady inspect things regularly?

2.		
Context	Reference	Item
Singular	Singular	Statistics show that every 11 seconds a man is mugged in New York City. This man regularly appears in the police station. The police take any incident reported very seriously.
Singular	Plural	Statistics show that every 11 seconds a man is mugged in New York City. These men regularly appear in the police station. The police take any incident reported very seriously.
Plural	Singular	CCTV footage showed that everyday a well-known mad man runs naked round the office. This man regularly appears in the police station. The police take any incident reported very seriously.
Plural	Plural	CCTV footage showed that everyday a well-known mad man runs naked round the office. These men regularly appear in the police station. The police take any incident reported very seriously.

3.		
Context	Reference	Item
Singular	Singular	Every school pupil in the country has their attendance marked on a register. This register is routinely kept in a very safe place. Pupils receive a certificate for good attendance at the end of the year.

Singular	Plural	Every school pupil in the country has their attendance marked on a register. These registers are routinely kept in a very safe place. Pupils receive a certificate for good attendance at the end of the year.
Plural	Singular	Every student in the class is listed on a register. This register is routinely kept in a very safe place. Pupils receive a certificate for good attendance at the end of the year.
Plural	Plural	Every student in the class is listed on a register. These registers are routinely kept in a very safe place. Pupils receive a certificate for good attendance at the end of the year.
Question		Are some pupils not on a register?

4.		
Context	Reference	Item
Singular	Singular	Every member of the congregation at the church said a private prayer. This prayer was usually at the end of the service. Attendees enjoyed the service.
Singular	Plural	Every member of the congregation at the church said a private prayer. These prayers were usually at the end of the service. Attendees enjoyed the service.
Plural	Singular	Every member of the congregation at the church listened to a sermon on Sunday. This sermon was usually at the end of the service. Attendees enjoyed the service.
Plural	Plural	Every member of the congregation at the church listened to a sermon on Sunday. These sermons were usually at the end of the service. Attendees enjoyed the service.
Question		Did the attendees not enjoy the service?

5.		
Context	Reference	Item
Singular	Singular	Every manager at the head office has a company car. This car regularly gets cleaned by local contractors. The head office accommodates staff as well as they can.
Singular	Plural	Every manager at the head office has a company car. These cars regularly get cleaned by local contractors. The head office accommodates staff as well as they can.
Plural	Singular	Every manager at the head office crowded into a boardroom. This boardroom regularly gets cleaned by local contractors. The head office accommodates staff as well as they can.
Plural	Plural	Every manager at the head office crowded into a boardroom. These boardrooms regularly get cleaned by local contractors. The head office accommodates staff as well as they can.

6.		
Context	Reference	Item
Singular	Singular	Every conservative MP has a loyal secretary. This secretary always remains polite at work. There are many requests from the general public which can make it challenging.

Singular	Plural	Every conservative MP has a loyal secretary. These secretaries always remain polite at work. There are many requests from the general public which can make it challenging.
Plural	Singular	Every visitor entering the small business reports to a receptionist. This receptionist always remains polite at work. There are many requests from the general public which can make it challenging.
Plural	Plural	Every visitor entering the small business reports to a receptionist. These receptionists always remain polite at work. There are many requests from the general public which can make it challenging.

7.		
Context	Reference	Item
Singular	Singular	Every new house in the estate has a 'For Sale' sign. This sign is always noticed by potential buyers. Buyers are quite pedantic when choosing a house to purchase.
Singular	Plural	Every new house in the estate has a 'For Sale' sign. These signs are always noticed by potential buyers. Buyers are quite pedantic when choosing a house to purchase.
Plural	Singular	Every new house in the street is in the shadow of a cathedral. This cathedral is always noticed by potential buyers. Buyers are quite pedantic when choosing a house to purchase.
Plural	Plural	Every new house in the street is in the shadow of a cathedral. These cathedrals are always noticed by potential buyers. Buyers are quite pedantic when choosing a house to purchase.

8.		
Context	Reference	Item
Singular	Singular	During the Second World War every soldier had a gun. This gun tragically resulted in numerous deaths. The loss of lives that occurred will always be remembered.
Singular	Plural	During the Second World War every soldier had a gun. These guns tragically resulted in numerous deaths. The loss of lives that occurred will always be remembered.
Plural	Singular	From 1914-1918 every soldier fought in a war. This war tragically resulted in numerous deaths. The loss of lives that occurred will always be remembered.
Plural	Plural	From 1914-1918 every soldier fought in a war. These wars tragically resulted in numerous deaths. The loss of lives that occurred will always be remembered.

9.		
Context	Reference	Item
Singular	Singular	Every astronomer records their observations in a diary. This diary successfully managed to get media attention. The media were very interested in this piece of news.
Singular	Plural	Every astronomer records their observations in a diary. These diaries successfully managed to get media attention. The media were very interested in this piece of news.

Plural	Singular	Every astronomer watched an eclipse on Sunday. This eclipse successfully managed to get media attention. The media were very interested in this piece of news.
Plural	Plural	Every astronomer watched an eclipse on Sunday. These eclipses successfully managed to get media attention. The media were very interested in this piece of news.
Question		Were the media uninterested?

10.		
Context	Reference	Item
Singular	Singular	Every employee in the office ate a sandwich at lunchtime. This sandwich naturally stopped people from working. The hour went by very fast but there was lots more work to be done.
Singular	Plural	Every employee in the office ate a sandwich at lunchtime. These sandwiches naturally stopped people from working. The hour went by very fast but there was lots more work to be done.
Plural	Singular	Every employee in the office heard a fire alarm. This alarm naturally stopped people from working. The hour went by very fast but there was lots more work to be done.
Plural	Plural	Every employee in the office heard a fire alarm. These alarms naturally stopped people from working. The hour went by very fast but there was lots more work to be done.

11.		
Context	Reference	Item
Singular	Singular	Every customer at Lloyd’s bank has a bank account. This account instantly attracted the local bank robber. It was not long before the bank robber was caught by the police.
Singular	Plural	Every customer at Lloyd’s bank has a bank account. These accounts instantly attracted the local bank robber. It was not long before the bank robber was caught by the police.
Plural	Singular	Every customer at Lloyd’s bank witnessed an explosion. This explosion instantly attracted the local bank robber. It was not long before the bank robber was caught by the police.
Plural	Plural	Every customer at Lloyd’s bank witnessed an explosion. These explosions instantly attracted the local bank robber. It was not long before the bank robber was caught by the police.

12.		
Context	Reference	Item
Singular	Singular	Every rice-grower has a makeshift wooden cart. This cart always requires some effort to make it work. Farm work can be very labour intensive but also very rewarding.
Singular	Plural	Every rice-grower has a makeshift wooden cart. These carts always require some effort to make them work. Farm work can be very labour intensive but also very rewarding.

Plural	Singular	Every dairyman follows a new government guideline on milk yields. This guideline always requires some effort to make it work. Farm work can be very labour intensive but also very rewarding.
Plural	Plural	Every dairyman follows a new government guideline on milk yields. These guidelines always require some effort to make them work. Farm work can be very labour intensive but also very rewarding.

13.		
Context	Reference	Item
Singular	Singular	Every ten seconds a woman gives birth. This woman understandably feels tired when she is finished. Time to relax is very important after doing something difficult like this.
Singular	Plural	Every ten seconds a woman gives birth. These women understandably feel tired when they are finished. Time to relax is very important after doing something like this.
Plural	Singular	Every week a local woman cleans my house. This woman understandably feels tired when she is finished. Time to relax is very important after doing something like this.
Plural	Plural	Every week a local woman cleans my house. These women understandably feel tired when they are finished. Time to relax is very important after doing something like this.
Question		Is taking time to relax very important after doing something like this?

14.		
Context	Reference	Item
Singular	Singular	Every week a team of climbers tackles Mount Everest. This team always unwinds when they have finished. They usually like to go to the pub for a well-deserved rest.
Singular	Plural	Every week a team of climbers tackles Mount Everest. These teams always unwind when they have finished. They usually like to go to the pub for a well-deserved rest.
Plural	Singular	Every year a committed team run the marathon for their charity. This team always unwinds when they have finished. They usually like to go to the pub for a well-deserved rest.
Plural	Plural	Every year a committed team run the marathon for their charity. These teams always unwind when they have finished. They usually like to go to the pub for a well-deserved rest.

15.		
Context	Reference	Item
Singular	Singular	Every ten minutes a person signs up for broadband internet. This person sadly has little knowledge of computers. Computers can be very easy to use when tasks are not too technical.
Singular	Plural	Every ten minutes a person signs up for broadband internet. These people sadly have little knowledge of computers. Computers can be easy to use when tasks are not too technical.

Plural	Singular	Every day an overworked technician gets told that the network is down. This technician sadly has little knowledge of computers. Computers can be easy to use when tasks are not too technical.
Plural	Plural	Every day an overworked technician gets told that the network is down. These technicians sadly have little knowledge of computers. Computers can be easy to use when tasks are not too technical.

16.		
Context	Reference	Item
Singular	Singular	Every time I cross the road a cyclist skips the red light. This cyclist usually gets to work on time. The workplace is not understanding if an employee is late.
Singular	Plural	Every time I cross the road a cyclist skips the red light. These cyclists usually get to work on time. The workplace is not understanding if an employee is late.
Plural	Singular	Every morning a friendly cyclist waves at me from the street. This cyclist usually gets to work on time. The workplace is not understanding if an employee is late.
Plural	Plural	Every morning a friendly cyclist waves at me from the street. These cyclists usually get to work on time. The workplace is not understanding if an employee is late.

17.		
Context	Reference	Item
Singular	Singular	Every day a soldier provides a selfless service in the war. This soldier persistently fights alongside the unit. This determination is admirable to many other soldiers.
Singular	Plural	Every day a soldier provides a selfless service in the war. These soldiers persistently fight alongside the unit. This determination is admirable to many other soldiers.
Plural	Singular	Every few miles a lost soldier stopped to ask for directions. This soldier persistently fights alongside the unit. This determination is admirable to many other soldiers.
Plural	Plural	Every few miles a lost soldier stopped to ask for directions. These soldiers persistently fight alongside the unit. This determination is admirable to many other soldiers.

18.		
Context	Reference	Item
Singular	Singular	Every driver drives a bus through the centre of the town. This bus is relatively old and should be replaced. An investment is required which the local council cannot afford.
Singular	Plural	Every driver drives a bus through the centre of the town. These buses are relatively old and should be replaced. An investment is required which the local council cannot afford.
Plural	Singular	Every child in the small village takes a bus to the High School in town. This bus is relatively old and should be replaced. An investment is required which the local council cannot afford.

Plural	Plural	Every child in the small village takes a bus to the High School in town. These buses are relatively old and should be replaced. An investment is required which the local council cannot afford.

19.		
Context	Reference	Item
Singular	Singular	Every delegate at the large international conference went to a local restaurant. This restaurant always had a wide selection of items. All the menu items tasted delicious.
Singular	Plural	Every delegate at the large international conference went to a local restaurant. These restaurants always had a wide selection of items. All the menu items tasted delicious.
Plural	Singular	Every secretary from the small firm went to a local restaurant. This restaurant always had a wide selection of items. All the menu items tasted delicious.
Plural	Plural	Every secretary from the small firm went to a local restaurant. These restaurants always had a wide selection of items. All the menu items tasted delicious.

20.		
Context	Reference	Item
Singular	Singular	Every school in Scotland stores their student details on a secure database. This database usually contains accurate information. Teachers find it useful to access the database online.
Singular	Plural	Every school in Scotland stores their student details on a secure database. These databases usually contain accurate information. Teachers find it useful to access the database online.
Plural	Singular	Every pupil in the class' name was entered into a database. This database usually contains accurate information. Teachers find it useful to access the database online.
Plural	Plural	very pupil in the class' name was entered into a database. These databases usually contain accurate information. Teachers find it useful to access the database online.
Question		Can the databases be accessed online?

21.		
Context	Reference	Item
Singular	Singular	Every athlete in the Olympics has had access to a personal trainer. This trainer quickly became fully booked. People were very excited to see the event.
Singular	Plural	Every athlete in the Olympics has had access to a personal trainer. These trainers quickly became fully booked. People were very excited to see the event.
Plural	Singular	Every runner in the race was told how to find a nearby stadium. This stadium quickly became fully booked. People were very excited to see the event.

Plural	Plural	Every runner in the race was told how to find a nearby stadium. These stadiums quickly became fully booked. People were very excited to see the event.

22.		
Context	Reference	Item
Singular	Singular	Every immigrant chose to work in a late-night supermarket. This supermarket always hires many staff over the summer. The workers are usually made permanent after their probation period.
Singular	Plural	Every immigrant chose to work in a late-night supermarket. These supermarkets always hire many staff over the summer. The workers are usually made permanent after their probation period.
Plural	Singular	Every one of Sarah's close friends decided to work in a late-night supermarket. This supermarket always hires many staff over the summer. The workers are usually made permanent after their probation period.
Plural	Plural	Every one of Sarah's close friends decided to work in a late-night supermarket. These supermarkets always hire many staff over the summer. The workers are usually made permanent after their probation period.
Question		Do the supermarkets lay off workers over the summer?

23.		
Context	Reference	Item
Singular	Singular	Every library user spent some time looking for a book. This book was typically about an interesting topic. It can be difficult to find time to relax and read.
Singular	Plural	Every library user spent some time looking for a book. These books were typically about an interesting topic. It can be difficult to find time to relax and read.
Plural	Singular	Every person in the book club agreed to read a book. This book was typically about an interesting topic. It can be difficult to find time to relax and read.
Plural	Plural	Every person in the book club agreed to read a book. These books were typically about an interesting topic. It can be difficult to find time to relax and read.

24.		
Context	Reference	Item
Singular	Singular	Every newcomer to the city took time to find a church. This church was very small and not well-known. The locals were very friendly and eager to help.
Singular	Plural	Every newcomer to the city took time to find a church. These churches were very small and not well-known. The locals were very friendly and eager to help.
Plural	Singular	Every wedding guest spent two hours searching for a church. This church was very small and not well-known. The locals were very friendly and eager to help.

Plural	Plural	Every wedding guest spent two hours searching for a church. These churches were very small and not well-known. The locals were very friendly and eager to help.

Appendix F: Full list of filler stimuli for Experiment 2 and 4.

No.	Item	Question
1	The woods were quite dangerous. As the eagle watched the birds, a mother who was very loving protected her new born. After some time, the eagle flew off, and the birds were safe.	Were the woods quite dangerous?
2	The medics had just started their shift. While the nurse was dressing a wound, a patient who was very ill was carefully being prepared for surgery. This was the first surgery the medics would watch	Had the medics just started their shift?
3	The bookshop looked untidy and lots of customers were waiting to pay. While the owner hunted for his glasses a pile of books suddenly fell on the floor. The customers were very understanding.	Were a lot of customers waiting in the book store?
4	The theatre organised a panto every Christmas. As the actor rehearsed his lines, another performance that was very long was almost at an end. The cast were excited to take to the stage on the first night of the panto.	Did the theatre organise a panto every Christmas?
5	Becky picked up a beautiful vase in the garden centre. As Becky ran to a staff member, passers-by were unsure of what had caused the accident. Becky needed a broom to clear the shattered vase pieces.	Did Becky pick up a beautiful vase?
6	The dinner had taken hours to prepare. While the man poured the gravy, some wine accidentally spilt over one of his guests. The guests laughed over the accident.	Did the man pour the gravy?
7	The markets had a variety of items. After the lady had studied the stalls she decided to stop at a jewellery stall. The lady saw a necklace that looked exactly like one her grandmother had.	Had the lady studied the stalls?
8	The man arrived home from his business trip a day early. As the man walked his dog a puppy that was quite friendly would not return when called by the owner. The puppy was not very well trained.	Was the man walking his dog?
9	The autumn leaves glistened brightly in the sunshine. The couple enjoyed their stroll that was quite long through the woods. They had wrapped up warm as it was very cold.	Did the couple enjoy their stroll?
10	The class had a test before the end of term. As the teacher tidied, the class that were quite young were	Did the teacher tidy?

	quietly whispering the answers to one another. Nobody wanted to fail the test.	
11	The film had been eagerly anticipated. While the men waited, the queue which was quite long, slowly shuffled forward. The men hoped they would see the pre-screening that evening.	Did the men hope to see the pre-screening?
12	It was a busy day in the office. As the receptionist rang a customer, the fire alarm that was extremely loud rang throughout the building. Test fire drills were scheduled to ring sporadically during the morning.	Was the fire drill scheduled to ring sporadically during the morning?
13	The lady was excited to go on holiday. While the pilot was flying, the plane which was quite large encountered turbulence and the captain had to land abruptly. The lady disliked flying.	Did the lady dislike flying?
14	The passengers boarded the congested train. As a man squeezed into his seat, another passenger who was quite scary began to quarrel over the reserved seat. The disagreement was resolved by the ticket collector.	Was the disagreement resolved by the ticket collector?
15	The Easter egg hunt was held every year. The child hunted for the golden egg that was quite hard to find while some kids sat eating all the chocolate. The parents were engrossed in other church activities.	Were the parents engrossed in other church activities?
16	The congestion on the road was extremely bad on Monday morning. While the lady was driving, a motorbike that was very loud stopped abruptly in front of her. The lady had to perform an emergency stop to avoid a collision.	Did the lady perform an emergency stop?
17	The plane had been delayed. As the man waited at arrivals many people were complaining about the delays at the airport. His parents eventually arrived hours later.	Was the man waiting at arrivals?
18	The tennis match had lasted for hours. As both players waited to be interviewed, the crowd which were very noisy chanted the winner’s name loudly. The match point had been very tense.	Had the tennis match lasted for hours?
19	Some friends went bowling together. During the third match, which was very tense, the friend with the least experience surprisingly won the game. The friends celebrated by buying their friend a drink.	Did some friends go bowling together?
20	The newly uncovered diary was thoroughly inspected. The historian thought about how the information could be cascaded. The diary was likely to be displayed in the museum and discussed on the radio.	Was the diary likely to be displayed in the museum?

21	Most students had arrived at school on time for the theme park trip. The coaches waited ready for the students to board. This would be the last trip for the students this year.	Did the coaches wait for the students to board?
22	Some shop owners in the town had many employees. The employees could become redundant as new working laws had been implemented by the government. The community hoped many jobs would not be lost.	Could the employees become redundant?
23	All of the Queen's hats are handmade. Designers hope to be able to design a suitable hat for the Queen. Blue hats are the Queen's favourite.	Do designers want to design a hat for the Queen?
24	All spies have many secrets to keep. Spies can find their job difficult. It can be difficult for spies to speak to their friends and family about their life.	Can spies find their job difficult?
25	Some students struggle with money at university. The government plan to freeze the rate of tuition fees. Some students feel the government are not doing enough.	Are the government planning on freezing the rate of tuition fees?
26	Most paintings were bought at the exhibition. The attendees were impressed with the work created. The painters were overwhelmed by the reaction that they received.	Were attendees impressed with the work?
27	Many people enjoy listening to music. Listening to different types of music genres is common. Some people may even go to concerts to hear music live.	Is listening to different types of music genres common?
28	Most billionaires are extremely hard-working. Wealthy people don't like to talk about their fortune. However, people often want to know how they became successful.	Are most billionaires extremely hard-working?
29	Lots of people dismiss the power of fashion. Some people believe fashion has the ability to change people's behaviour and attitudes in society. Designers work hard to make statements using fashion.	Do lots of people dismiss the power of fashion?
30	Many young people like to take selfies. Some social historians see selfies as evolved self-portraits. Others believe it is narcissistic.	Do many young people like to take selfies?
31	Everyone should eat five fruit and or vegetables a day. Some people can find it difficult to meet this goal. As obesity is on the rise a good diet and regular exercise is important to stay healthy.	Should everyone eat five fruit and or vegetables a day?
32	All doctors take an oath to help their patients. Patients are often thankful for the care that they receive. It is a rewarding job when a patient's health improves.	Do all doctors take an oath to help their patients?

33	Most driving tests are difficult to pass. Learners can find it hard if they are very nervous. Driving instructors try to prepare students as best they can.	Are most driving tests difficult to pass?
34	Many sports car racers enjoy driving very fast. Racing car drivers are motivated to race when they win a trophy. Sports car racing can be a dangerous sport.	Can sports car racing be a dangerous sport?
35	Most smart phones can be very convenient devices to help individuals communicate. People have a mobile phone which they can use to search for information. Phones can be disruptive especially when driving.	Can phones be disruptive especially when driving?
36	Many interview candidates are nervous when being interviewed for a job. There are techniques on how to cope with nerves in an interview. The interviewing panel enjoy conducting interviews.	Do the interview panel enjoy conducting interviews?
37	Types of bad habits are overspending, procrastinating and biting finger nails. Some people have good will power to combat their bad habits. It can take some time but bad habits can improve.	Do bad habits always get worse?
38	Email is a quick and easy way to write a lot of information to someone. Some people receive a lot of spam emails. Email is also a low-cost form of communication.	Is email an expensive form of communication?
39	Every lecture is boring! Many lecturers try to make lectures interesting for students. Some students fall asleep during their lectures.	Do all students stay awake during lectures?
40	Most news articles are correct. Journalists spend a lot of time writing interesting and well-researched articles. Newspapers are being read less frequently.	Are people reading newspapers more frequently?

Appendix G: Experiment 2 – Syntactic ambiguity post hoc analyses

Linear and binomial mixed-effects models post hocs at the pre-critical region.

	Pre-critical region			
	Estimate	SE	t/z value	
First pass reading times				
Intercept	416.307	12.741	32.68	***
Long sentence: Ambiguous v Unambiguous	24.749	4.161	5.95	***
Short sentence: Ambiguous v Unambiguous	38.409	4.447	8.64	***
Regression path reading times				
Intercept	645.837	13.41	48.16	***
Long sentence: Ambiguous v Unambiguous	137.176	11.723	11.70	***
Short sentence: Ambiguous v Unambiguous	171.679	9.178	18.71	***
Intercept	635.727	10.432	60.94	***
Long sentence: Young v Older adults	-31.226	9.340	-3.34	***
Short sentence: Young v Older adults	0.704	7.690	0.09	
Intercept	604.291	8.550	70.68	***
Ambiguous sentence: Young v Older adults	-29.867	7.539	-3.96	***
Unambiguous sentence: Young v Older adults	-11.098	6.691	-1.66	
Intercept	602.439	10.84	55.58	***
Ambiguous, long sentence: Young v Older adults	-52.248	7.633	-6.85	***
Unambiguous, long sentence: Young v Older adults	-14.735	8.629	-1.71	
Ambiguous, short sentence: Young v Older adults	-18.991	9.904	-1.92	
Unambiguous, short sentence: Young v Older adults	-10.062	9.586	-1.05	

Regressions Out				
Intercept	-1.937	0.146	-13.26	***
Long sentence: Young v Older adults	-0.233	0.103	-2.26	*
Short sentence: Young v Older adults	-0.010	0.102	-0.10	
Intercept	-1.754	0.145	-12.08	***
Long sentence: Ambiguous v Unambiguous	0.336	0.086	3.92	***
Short sentence: Ambiguous v Unambiguous	0.574	0.087	6.59	***
Probability of skipping				
Intercept	3.987	0.386	10.33	***
Long sentence: Ambiguous v Unambiguous	-0.066	0.140	-0.47	
Short sentence: Ambiguous v Unambiguous	0.461	0.141	3.26	**
Total reading times				
Intercept	764.167	6.432	118.80	***
Long sentence: Ambiguous v Unambiguous	155.795	5.968	26.11	***
Short sentence: Ambiguous v Unambiguous	168.762	6.045	27.92	***
Intercept	770.751	10.086	76.42	***
Long sentence: Young v Older adults	2.099	26.469	0.079	
Short sentence: Young v Older adults	43.553	16.459	2.646	**
Intercept	617.512	8.949	69.00	***
Ambiguous sentence: Young v Older adults	13.051	8.015	1.63	
Unambiguous sentence: Young v Older adults	29.607	8.048	3.68	***

Linear and binomial mixed-effects models post hocs at the critical region.

Critical region				
	Estimate	SE	t/z value	
First pass reading times				
Intercept	436.609	7.862	55.53	***
Older adults, ambiguous sentence: long v short	-10.213	5.722	-1.79	
Older adults, unambiguous sentence: long v short	-0.0585	5.386	-0.01	
Young adults, ambiguous sentence: long v short	-1.700	5.507	-0.30	
Young adults, unambiguous sentence: long v short	-10.151	4.875	-2.08	*
Regression path reading times				
Intercept	775.229	10.372	74.75	***
Ambiguous sentence: Long v Short	166.560	8.418	19.79	***
Unambiguous sentence: Long v Short	107.311	7.287	14.73	***
Intercept	702.535	6.921	101.50	***
Long sentence: Young v Older adults	3.457	6.286	0.55	
Short sentence: Young v Older adults	37.405	6.612	5.66	***
Intercept	702.803	7.014	100.20	***
Ambiguous sentence: Young v Older adults	-21.570	9.781	-2.21	*
Unambiguous sentence: Young v Older adults	9.1790	8.029	1.14	
Total reading times				
Intercept	722.356	9.968	72.47	***
Long sentence: Young v Older adults	-25.366	11.701	-2.17	*

Short sentence: Young v Older adults	-15.321	9.307	-1.65	
Intercept	709.520	10.910	65.06	***
Ambiguous sentence: Young v Older adults	36.470	92.470	0.39	
Unambiguous sentence: Young v Older adults	57.420	28.240	2.03	*

Linear and binomial mixed-effects models post hocs at the critical region.

	Post-critical region			
	Estimate	SE	t/z value	
First pass reading times				
Intercept	530.364	10.858	48.84	***
Long sentence: Young v Older adults	21.582	9.172	2.35	*
Short sentence: Young v Older adults	33.218	7.816	4.25	***
Intercept	517.404	7.852	65.89	***
Ambiguous, long sentence: Young v Older adults	-25.546	5.477	-4.66	***
Unambiguous, long sentence: Young v Older adults	-21.417	6.090	-3.52	***
Ambiguous, short sentence: Young v Older adults	-5.014	5.041	-1.00	
Unambiguous, short sentence: Young v Older adults	-18.55	5.651	-3.28	**
Regression path reading times				
Intercept	1033.993	10.247	100.90	
Long sentence: Young v Older adults	-39.893	8.039	-4.96	
Short sentence: Young v Older adults	8.435	12.270	0.69	
Intercept	952.241	8.052	118.26	

Ambiguous sentence: Young v Older adults	109.164	8.130	13.43	
Unambiguous sentence: Young v Older adults	72.128	8.392	8.60	
Total reading times				
Intercept	724.776	6.944	104.38	***
Long sentence: Young v Older adults	34.736	7.873	4.41	***
Short sentence: Young v Older adults	45.903	7.616	6.03	***
Intercept	710.950	9.847	72.20	***
Older adults, ambiguous sentence: long v short	-37.585	6.478	-5.80	***
Older adults, unambiguous sentence: long v short	-23.857	7.460	-3.20	**
Young adults, ambiguous sentence: long v short	-4.890	7.372	-0.66	
Young adults, unambiguous sentence: long v short	-12.365	6.314	-1.96	

Appendix H: Experiment 4 – Semantic ambiguity post-hoc tests analyses

	Critical region			
	Estimate	SE	t/z value	
First Fixation Duration times				
Intercept	232.415	5.711	40.87	***
Plural Context, Singular reference v Plural Context, Plural reference contrast	2.333	2.234	1.05	
Singular Context, Singular reference v Singular Context, Plural reference	6.456	2.313	2.79	**
Regression-path reading times				
Intercept	542.432	8.764	61.89	***
OA: Singular v plural context	-19.105	9.177	-2.08	*
YA: Singular v plural context	8.573	9.091	0.94	
Intercept	554.318	10.188	54.41	***
OA: Singular v plural reference	16.393	8.639	1.90	.
YA: Singular v plural reference	39.838	10.157	3.92	***
Total reading times				
Intercept	592.375	9.160	64.67	***
Plural Context, Plural reference v Plural Context, Singular reference contrast	7.755	8.331	0.93	
Singular Context, Plural reference v Singular Context, Singular reference contrast	33.548	8.954	3.75	***
Intercept	587.921	10.179	57.76	***

Plural Context, Plural reference: YA v OA	14.475	9.28	1.56	
Plural Context, Singular reference contrast: YA v OA	34.723	8.862	3.92	***
Singular Context, Plural reference: YA v OA	28.400	10.410	2.73	**
Singular Context, Singular reference contrast: YA v OA	19.703	9.072	2.17	*

	Post-critical region			
	Estimate	SE	t/z value	
First-Pass reading times				
Intercept	455.384	8.987	50.67	***
Plural Context, Singular reference v Plural Context, Plural reference contrast	11.416	4.418	2.58	**
Singular Context, Singular reference v Singular Context, Plural reference	2.713	4.556	0.60	
Intercept	454.728	9.981	45.559	
Plural reference contrast: YA v OA	12.457	7.846	1.59	
Singular reference contrast: YA v OA	23.212	7.906	2.94	**
Regression-path reading times				
Intercept	560.767	9.665	58.02	***
Plural Context, Singular reference v Plural Context, Plural reference contrast	24.063	6.771	3.554	***
Singular Context, Singular reference v Singular Context, Plural reference	5.279	7.570	0.70	
Total reading times				

Intercept	653.798	14.116	46.32	***
Plural Context, Plural reference v Plural Context, Singular reference contrast	32.910	7.348	4.48	***
Singular Context, Plural reference v Singular Context, Singular reference contrast	6.087	6.029	1.01	
Intercept	680.095	12.274	55.410	***
Plural Context: YA v OA	58.308	10.246	5.69	***
Plural Context: YA v OA	44.016	9.459	4.65	***
Intercept	676.345	8.78	77.03	***
Plural Context, Plural reference: YA v OA	41.882	8.736	4.79	***
Plural Context, Singular reference contrast: YA v OA	48.257	9.644	5.00	***
Singular Context, Plural reference: YA v OA	32.035	8.793	3.64	***
Singular Context, Singular reference contrast: YA v OA	34.375	7.889	4.36	***