Situational Ambiguity and Ageing : Navigating the ambiguous world of underspecified situations in current and later life

Calvin Laursen
A thesis submitted in partial fulfillment of the requirements of Bournemouth University for the degree of Doctor of Philosophy 2024

Copyright Statement

Abstract

Ambiguity detection, understanding if a text has more than one valid interpretation, in mental models, is not well understood. We developed a novel paradigm for investigating if, how, and when readers of texts that describe a situation detect ambiguity. We also investigated how texts which describe spatial situations are processed differently from texts that describe non-spatial situations and how ageing affects processing. My doctoral thesis presents a novel paradigm for investigating situational models, from paragraphs, which better resemble “real life”, by embedding premises into paragraphs. We found that ambiguity is hard for readers to detect but suggest that encouraging, or discouraging certain reasoning strategies we improved rates of ambiguity detection. We speculate that using certain reasoning strategies allows participants to create a more complete model from the premises which is more conducive to ambiguity detection. We did this by changing how the reasoner’s understanding of a model is probed during questioning. Specifically, we found that probing the “internal” part of the model significantly increased ambiguity detection. In line with prior research (e.g., Light, 1988; Radvansky et al.,1990) we found that that despite slower reading (Myerson et al., 1990) and processing (Copeland & Radvansky, 2007), older participants are just as capable as younger readers at creating mental models and detecting ambiguity. We suggest that older participants may use more holistic reasoning strategies, e.g.encoding the entire model rather than the two premise parts. Which has been suggested previously as a compensatory mechanism for lower working memory capacity (Copeland & Radvansky, 2007). Our findings lead us to conclude that situational ambiguity is harder to detect for both older and younger people and that this is highly dependent on strategies, as probing different different parts of the model yield different rates of ambiguity detection.

Thesis Structure

Chapters 2, 4 and 8 are included as distinct articles written for publication in peer-reviewed journals. In order to avoid repetition the introductions of Chapter 4, 6 and 8 do not repeat previously introduced information and focus on new elements. Published articles based on these chapters will contain this omitted information. Chapter 1 serves as an introductory. Chapter 4 is a study in its own right but serves more as a proof of concept as the study is underpowered. The linking chapters also discuss lines of investigation not included in the manuscripts. As such repetition between chapters is inevitable, this is particularly true for the introductions and especially when comparing the introduction to the submitted article and the general introductions. Each table and figure contained in this thesis is labelled in respect to its chapter and are marked with a number marking the chapter . A list of these figures and tables can be found in the table of content (Page 6).

Current publication status of papers within this thesis

Chapter 6 - “Exploring Ambiguity in Situational Models: Evidence from Reading Comprehension” has been split into three proposed papers of which, at the time of submission (28/07/2024), one is available in pre-print format while the other two are in preparation.

Table of Contents

Copyright Statement 2

Thesis Structure 4

Current publication status of papers within this thesis 5

Table of Contents 5

List of Figures and Tables 11

Acknowledgements 17

About this Thesis 18

Chapter 1: A general introduction to ambiguity and mental models. 20

What is ambiguity, why are we ambiguous and what are mental models? 20

The construction of a mental model. 28

Differences between this Thesis and prior mental model research 33

Ageing 36

This thesis 39

Abstract: 42

1.0 Introduction 43

1.1 Differences between mental model research and secondary research questions 48

2.0 Method 51

2.1 Participants 51

2.2 Materials and Design 51
2.3 Stimuli Details 54
2.4 Procedure 56
2.5 Analysis 57
3.0 Results: 57
3.1 Reading times of stimuli (Described situation) 58
3.2 Question reading time 60
3.3 Question accuracy 62
3.4 Likelihood to select “There is not enough information” 66
4.0 Discussion 70
5.0 Conclusion 76
Chapter 3: Implications of situational ambiguity and a general introduction to ageing 78
Reflecting on Chapter 2 78
Aging 83
Abstract: 88
Introduction: 89
Improving methodology: 89
Reading, eye-movements and ageing: 92
This Study 101
Participants 103
Experiment Design 104

Descriptive information about the stimuli 106
Apparatus 107
Procedure 108
Analysis 108
Results 110
Exclusion 110
Question accuracy 110
Question reading time 112
Risky reading 116
Skip rate 116
Regression probability 118
Fixation Duration and Number of Fixations 120
Discussion 122
Conclusion 127
Abstract 131
Introduction 132
Experiment 1 137
Transparency and Openness 137
Participants 138
Materials and Designs 139
Descriptive information about the stimuli 140

Procedure 142

Analysis 146

Results 147

Reading time of Stimuli 147

Question reading time 149

Question accuracy 153

Line by Line Reading Times 156

Model selection 159

Discussion of Experiment 1: 163

Experiment 2 167

2: Participants 168

2: Results 170

2: Reading time of Stimulus text 170

2: Question reading time 172

2: Question Accuracy 175

2: Line by Line Reading Time 179

2: Model selection 183

Discussion Experiment 2: 186

Internal/External Comparison 190

3: Question reading time 191

3: Question Accuracy 191

3: Model selection 194

3: Discussion of the internal/external comparison 197

Conclusion 202

Abstract 204

Reflections on Chapter 5: 205

1.0 Introduction: 207

1.1 This study 211

1.2 Participants 215

1.3 Materials and Design 215

1.4 Procedure 219

1.5 Analysis 224

1.6 Results 225

1.6.1 Reading time of Stimuli 225

1.6.2 Modulation of SRT by trial order 228

1.6.3 Question Reading time 232

1.6.4 Modulation of QRT by trial order 235

1.6.5. Question Accuracy 240

1.6.6 Modulation of Question Accuracy by trial order 244

247

1.6.7 Premise Line by line reading time 248

1.6.8 Premise Line by line re-reading and reading time by trial order 251

Discussion 254

Conclusion 261

Chapter 9: General Discussion and Conclusion 262

Future research 267

Conclusion 269

Chapter 9: General Discussion and Conclusion 270

Future research 275

Conclusion 277

Chapter 10: Conclusion 278

References 279

Appendix: 294

List of Figures and Tables

Table1: Possibilities of A and B’s locations and whether they are consistent with the premise 17

Figure 5a. An example stimulus: Item 3 in the configuration Unambiguous non-spatial first, Unambiguous spatial second. 47

Figure 5b. An example question for spatial Item 1, participants would select their answer using one of the answer buttons shown. 47

Table 1: 49

Table 2.0: Mean stimulus reading times in ms, mean total QRT, and mean comprehension question reading time(SD in parenthesis). 52

Table 2.1: 53

Table 3: LMM Results for QRT log(ms) as a function of ambiguity and stimulus type (non-spatial/spatial). 55

Figure 7: Two-way interaction between Ambiguity and Question Type. The values were extracted using the “effects” R package v.4.1 (Fox & Hong, 2009) and Plotted using ggplot package for R (Wickham 2016). 56
Table 4: 57
Table 5: 57
Figure 8:Two-way interaction between Ambiguity and Type influencing question accuracy. The values were extracted using the “effects” R package v.4.1 (Fox & Hong, 2009) and Plotted using ggplot package for R (Wickham 2016). 59
Table 6. 62
Figure 9. Predictive model showing the interaction between Spacial/non-spatial, Ambiguous/unambiguous, and QRT, influencing selecting “There is not enough information”, fitted using GAMM smooths 63
Figure 1a: Unambiguous non-spatial stimulus (premises highlighted in yellow) 99
Figure 1b:Example of a model comprehension question 99
Table 1: Means table showing readability metrics 100
Table 2.0: Mean question accuracy by condition (SD in parentheses) 105
Table 2.1:GLMM Results for question accuracy as a function of ST,Ambiguity and AgeGroup. 106
Table 3:Mean question reading time (ms) by condition (SD in parentheses) 107
Table 4:Results of a linear mixed model of Question reading time as function of ST,Ambiguity and AgeGroup. 108
Table 6: Results of a linear mixed model of Skip frequency as function of Ambiguity, Premise word and AgeGroup. 110
Table 7: GLMM results table for regression likelihood as a function of Age,ST, Ambiguity and premise word. 113
Table 8: LMM results showing fixation duration (top) and number of fixations (bottom) as a factor of Age and Premise word 115
Table 1: Mean readability statistics for the stimuli and questions (SD in parenthesis). 135
Figure 1: Stimulus as presented on screen with the first line of text unmasked 137
Figure 2a: Question screen showing a spatial question 140
Figure 2b: Question screen showing a non-spatial Question 140
Table 2:Mean stimulus reading times in ms (SD in parenthesis). 142
Table 3:Mean comprehension and situational question reading times in ms (SD in parenthesis) . 144
Table 4:LMM Results for question reading time (QRT) log(ms) as a function of the presence of ambiguity, Age and stimulus type (ST: non-Spatial/spatial). 145
Figure 3: LMM estimates for question reading time (QRT) log(ms) as a function of the presence of ambiguity, Age and stimulus type (ST: non-Spatial/spatial). Means shown by red bar 146
Table 5:Mean question accuracy by ST, Ambiguity and Age (SD in parenthesis). Note: General comprehension questions were the same across all conditions. 148
Table 6:GLMM Results for question accuracy as a function of Ambiguity, ST (non-spatial/spatial) and Age. 149
Table 7:LMM results for line reading time as a function of premise, and Age. 151
Table 8:GLMM results for re-reading as a function of premise, Age and Ambiguity. 152
Table 9:Mean likelihood to select premise 1 or premise 2, between premise 1 and 2, if ambiguity is undetected, by ST(spatial/non-spatial) and age group. 154

Table 10:LMM estimates of likelihood to prefer one model over another as a factor of ST(spatial/non-spatial),Premise (Premise 1 = 0, Premise 2 = 1), and Age Group. 155
Figure 4: Likelihood of model selection by Premise 1 or 2, Ambiguity detection or the Anchor 156
Table 9: Descriptive information of questions probing the “internal” part of the model used in Experiment 2 163
Table 10:Mean stimulus reading times in ms (SD in parenthesis). 165
Table 11: Mean Question reading times in ms and mean comprehension question reading time(SD in parenthesis). 167
Table 11: LMM Results for question reading time (QRT) log(ms) as a function of Age, the presence of ambiguity, and stimulus type (ST; non-Spatial/spatial). 168
Table12:Mean question accuracy of older readers by ST and the presence of ambiguity and Age (SD in parenthesis). 170
Table 13:GLMM Results for question accuracy as a function of the presence of ambiguity, ST (non-Spatial/spatial) and Age. 171
Table 14:LMM results for line reading time as a function of premise, and Age. 173
Table 15:GLMM results for re-reading as a function of premise, and Age. 174
Table 16:Mean likelihood to select premise 1 or premise 2, between premise 1 and 2, if ambiguity is undetected, by ST(spatial/non-spatial) and age group. 176
Table 17:LMM estimates of likelihood to prefer one model over another as a factor of ST(spatial/non-spatial),Premise (Premise 1 = 0, Premise 2 = 1), and Age Group. 177
Figure 5: Selection likelihood violin plot showing the difference in participants selecting a model based on Premise 1, Premise 2, the anchor or answering the question correctly. 178
Table 17: GLMM results for Question accuracy as a function of Persp (Internal/External) interacting with Age group, ST, and the presence of ambiguity, between the two groups. 184
Figure 6: Differences in questions accuracy between Internal (Experiment 2) and External (Experiment 1) Questions. 185
Table 18: Comparative likelihood to select a model based on one premise over another as a factor of which premise had a higher selection likelihood (Premise 1vs Premise 2), Internal/External probing questions and AgeGroup 188
Figure 7:Violin plot showing preference to select one model over another, Internal vs. External question types. 189
Table 1: Table showing mean readability metrics of the stimuli and questions. (A full set of stimuli and questions can be found via the OSF link) 211
Figure 1: Stimulus as presented on screen with the first line of text unmasked 213
Figure 2a: Question screen showing an Anchor question 215
Figure 2b: Question screen showing an Ambiguous Question 215
Table 2. LMM results for SRT as a factor of ST,BT and QT the random factors included were item,subject and age. 219
Figure 3: Predictive GLMM model plot for stimulus reading time, showing interaction between ST,BT and QT. 221
Figure 4: Predictive GAMM model showing Trial reading by (smoothed) BT&QT over trials. The thick black line indicates the block change. 223
Table 4. LMM results for SRT as a factor of ST,BT and QT. Including only trials from 20-30, the random factors included were item,subject and age. 224
Figure 5: Block transition change in trial reading time for Ambiguous questions first (Red) and Unambiguous first (Blue), Block number 1 refers to 5 trials prior to block transition and Block number 2 refers to 5 trials after block transition. 225
Table 5. Means Table of Question reading time 226

Table 6.LMM results Table showing Log (Question reading time) as a factor of BT,QT and ST 227
Figure 6: Predictive plot based on LMM results shown in Table 6, Log (Question reading time) as a factor of BT,QT and ST 228
Figure 7: Predictive GAMM model showing Question reading time by (smoothed) BT & QT over trials. 230
Table 6.LMM results Table showing question reading time of trials from 20-30 as a function of BT and Block number 231
Figure 8. Block transition change in question reading time for Ambiguous questions first (Red) and Unambiguous first (Blue), Block number 1 refers to 5 trials prior to block transition and Block number 2 refers to 5 trials after block transition. 232
Table 7: Means Table of Question accuracy 234
Table 8:GLMM results Table showing question accuracy (%) as a function of BT,QT and ST 235
Figure 9:Predictive boxplot showing the interaction between block type, question type and subject type. 236
Figure 10: Predictive GAMM model showing question accuracy (%) as a function of (smoothed) ST,BT,QT over trials. 238
Table 9:GLMM results for Question accuracy of trials from 20-30 as a function of Block number and BT. 239
Figure 11:Block transition change in question accuray for Ambiguous questions first (Red) and Unambiguous first (Blue), Block number 1 refers to 5 trials prior to block transition and Block number 2 refers to 5 trials after block transition. 240
Table 10. LMM results for Premise line reading time in (log) ms as a factor of Block type and Premise 242
Table 11. GLMM results for Premise line re-reading likelihood as a factor of Block type and Premise. 243
Figure 7: Predictive GAMM model of Premise line re-reading time by sequence 245
Figure 8: Predictive GAMM model for premise line re-reading likelihood by sequence 246

Acknowledgements

Over the course of this PhD and writing of this thesis many people were of great help to me. I would like to hereby thank anyone who listened to me talk about my research and asked me questions about it even though you may not have understood what it was I was talking about. Of the people who helped me some significantly stand out and I shall do my best to give them my thanks.

Firstly, to my supervisors Tim Slattery, Jan Wiener and Martin Vassilev who's guidance allowed me to continue throughout the Covid-19 Pandemic and were there to help every step of the way. You are patience and wisdom personified. Additional special thanks go to Jan Wiener and Martin Vassilev for putting up with me for as long as they have.

To my friend Victoria Nicholls, for helping me understand statistics and visualisations better.

To my friend Rachel Skinner, for helping me believe in myself.

To my partner Luka Morkyte, for being there for me every step of the way and for proof reading almost everything I have ever written. I promise to do the same for you. I do not know how I could have done any of this without you.

To my friend Claudia Tang, for teaching me to practise what I preach. I promise I will take some time off after this.

To my friend Catherine Griffiths for distracting me with renal electron microscopy pictures. I still don't know what it is I'm supposed to be looking at but I am glad you're there to tell me.

To my friend Fay Sweeting for being an emotional crutch when needed.

To all of my older participants, without whom this research would not have been possible.

Additional special thanks to:

Anyone who listened to me talk about my work which may have come across as the excited ramblings of a toddler.

The U3A

The ADRC

Bournemouth University

Nicotine, Caffeine and Sertraline.

About this Thesis

This thesis explores how situational ambiguity is interpreted by readers and how this differs between older and younger people. However, it was not the original research question of this thesis. To begin with, this thesis was supposed to investigate how the return sweep, an eye movement from the end of one line to the start of the next, changed as we aged. When work on this thesis began in 2019 it was understood that most if not all research relating to return sweeps would involve eye-tracking. However, when it became apparent that in-person eye-tracking would not be possible due to the COVID-19 pandemic, research shifted to what was originally a secondary, more exploratory aspect of the research.

The shift in focus is not too apparent in this thesis, however, chapter 6 was designed and written when it was once again possible to test face-to-face using an eye-tracker as eye-tracking remains a useful tool for understanding how text is interpreted. With that said, Chapter 6 also pays some homage to the original research question and gives some insight into the return sweep, despite this being somewhat out of place with the rest of the thesis.

The thesis explores how mental models, created from premises in a body of text are interpreted, and when these are ambiguous, i.e underspecified, if a reader can detect this ambiguity. The original methodology made it possible to relatively quickly shift the focus from eye-movements to parameters that made it possible to explore these questions in detail. Most research into how mental models are created and interpreted use relatively simple, single line texts. However, this thesis outlines a novel approach in exploring mental models through a more “naturalistic” reading task whereby the mental model is outlined through premises within a larger body of text. Moreover, our methodology also differed from prior research into mental models in that we asked participants to determine whether or not the situation described in the text was ambiguous, i.e we tested their ability to detect ambiguity. Furthermore, we investigated the influence of subject matter, comparing how spatial and non-spatial descriptions influence the construction and inference from mental models, and ambiguity therein. Lastly, as with the original concept of this thesis, we investigatedthe effects of ageing on ambiguity detection and how known reading processes may influence ambiguity detection.

Ambiguity detection, ageing, and subject matter differences in mental models are understudied. Therefore, it is important to note that we use results from ambiguity research from reading and discourse processing to interpret our results. This research highlights that humans use ambiguous language in our day to day lives and how this ambiguity is processed.

The full reproducible codes, raw data, individual chapters can be found on the OSF for this Thesis https://osf.io/rydeh/?view_only=b75558090201482da9f1ac27dc0fa023

Chapter 1: A general introduction to ambiguity and mental models.

What is ambiguity and what are mental models?

Originally, mental models were proposed as a programmatic basis for thinking (Craik, 1943). Mental models are thought to be pivotal to the understanding of discourse whereby processing verbal or written structures lead to an internal or mental model of the described situation akin to the result of perceiving or imagining the situation. These models are constructs used not only for language processing (Johnson-Laird 1983) but also for spatial reasoning (e.g. Vosgerau, 2006). Mental model theory posits that mental models are tools with which humans manipulate mental objects and arrays in order to draw conclusions and make inferences.

Human spatial reasoning is vital for the navigation of our world and in order to do this we construct mental models. We do this in order to locate, rotate and replace objects in specific orders (Ragni et al., 2007). Furthermore, spatial reasoning covers a wide range of situations from objects relating to one another to finding where one is in relation to other objects/subjects. For example a relational proposition commonly found in mental model research would be:

Example 1.0:

"The cup is to the left of the pen"

A stimulus, such as the premise above, allows a reasoner to make inferences about the location of the cup and the pen in relation to one another. Researchers of mental models (e.g., Goodwin & Johnson-Laird, 2005; Fangmeier et al., 2006) refer to "the pen" and "the cup" from the example above as "terms", however we will refer to these simply as "objects/subjects", as in this Thesis these "terms" refer to both objects and people.

Mental model theory (Johnson-Laird & Byrne, 1991) stipulates that people make inferences from an array they construct from given information. A reasoner, having read

Example 1.0, would therefore be able to conclude not only that the cup is to the left of the pen but also that the pen is to the right of the cup. Creating such a simple model poses no great difficulty for a reasoner (Goodwin & Johnson-Laird, 2005), but the more objects/subjects and their relationships described by the premises, the harder the model becomes to create and make inferences from. For example:

<u>Example 1.1:</u>

"Imagine **you** are standing at the entrance to a park, to your **right** is an information centre, to your **left** is a rubbish bin."

This information allows you to create a mental model of the spatial situation around you:

<u>Example 1.2:</u>

*" **Rubbish bin – You – information centre**"*

In this example the positions of all objects/subjects, in relation to one another, are known and a reasoner can easily create a mental model of the described situation. In general, mental model theory describes how people can translate verbal or written language into a mental representation from which to make inferences (Johnson-Laird & Byrne, 1991) and that these mental representations do not take the form of logical propositions (Byrne and Johnson-Laird, 2009). Rather mental models are constructed by envisioning the situation and alternative possibilities using information from perception or the comprehension of discourse/text.

A mental model, in contrast to logical propositions, is based on a set of fundamental assumptions, outlined by the premises provided and knowledge of the world (Byrne and Johnson-Laird, 2009). Former theories on reasoning imply that reasoning makes use of rules of inference that are similar to axiomatic set theory or calculus (O'Brian, 2009). For example;

"If A is on the left then B is on the right.

A.

Therefore, B is to the right of A."

Such a logical set immediately gives the reasoner a conclusion "A is left of B". However, using "pure" calculus like logic becomes more complex if a negative inference is used. Suppose that:

"If A is on the left, then B is on the right".

"B is not on the right".

It follows that:

"A is not on the left".

While this supposition allows us to determine that A is not on the left it does not allow for logical inferences beyond this determination. Even if we were to suppose that "B is on the left" we cannot infer any more information about this small model as such a statement directly contradicts the premise. Additionally, it begs the question "To the left/right of what?" as a reference point is absent. From the premise, we might (rightly) conclude that left and right are describing the relationship between A and B and that any supposition in relation to which side A and B are on simply implies a switch in position between A and B. The conclusion therefore would be that the premise is false. Formal rule theories (e.g., O'Brian, 2009) postulates *"that making inferences is a search for a proof of a conclusion, and that if* *no such proof is discovered then the inference is invalid"* (378, Byrne and Johnson-Laird, 2009). However, as the conclusion is that the premise is false any inferences one could make from it are therefore also false. Using calculus like logic to determine the order of objects/subjects within a model is not particularly useful.

Contrary to using logic to make inferences from a premise, a mental model allows the reasoner to "envisage" possibilities based on the premise. A possibility in which the conclusion does not hold true with the premise is rejected rather than invalidating the premise, while a conclusion which holds true to the premise is "accepted". Suppose again that:

"If A is on the left, then B is on the right".

It is possible to create multiple different outcomes as to where A and B are based on envisaged possibilities.

Table1: *Possibilities of A and B's locations and whether they are consistent with the premise*

A	**B**	**Consistency**
Left	Right	Consistent
Not Left	Right	Consistent
Right	Left	Consistent
Not Right	Left	Consistent
Not Left	Not Left	Consistent
Not Right	Not Right	Consistent
Left	Left	Inconsistent
Right	Right	Inconsistent

By envisioning only 8 of the possible relationships between A and B (*relationships such as "A under B" or "B not on the left, A behind B" are excluded*) it is possible to see that none of them necessarily invalidate the premise and that so long as the relationship is consistent with the premise inferences made are valid. It should be noted that if premises contain "If and only If" statements then suppositions are much easier to reason with, though this is not the focus of this thesis.

By contrast, if a premise lacks an "If" e.g., "*A **is** on the left, B **is** on the right"* then there is only one conclusion which does not invalidate the premise as premise and conclusion are one and the same. Negative conclusions also work with a "non-if" premise e.g., " B is not on the left". The only way in which a statement such as "A is on the left, B is on the right" can have multiple interpretations is by adding an additional premise which, while consistent with the other premise, leaves the conclusion open.

Returning to *Example 1.1*:*"Imagine **you** are standing at the entrance to a park, to your **right** is an information centre, to your **left** is a rubbish bin*." it is not possible to construct a table of possibilities like those in *Table 1* as the example lacks "if" statements. Rather the "*information centre"* is to your right and the *"rubbish bin"* is to your left. Any conclusions that do not hold true to those two premises are invalid and no inferences can be made from incorrect suppositions. However, changing one of the premises allows for the existence of multiple viable conclusions and situations in which more than one mental model is viable. Note also that "non-if" premises require at least three objects/subjects in relation to one another. By changing the second premise in *Example 1.1* to "*to your **right** is a rubbish bin"*, more than one consistent solution to the "left to right" order of the objects/subjects is possible.

If mental models are generated "bit by bit" (Johnson-Laird & Byrne, 1991), the second, changed, premise is "added" to the first "*to your **right** is an information centre"* and must now be integrated with the simple model created by the first premise. Because the second premise has been changed, the solution outlined in *Example 1.2* is now invalid and a new model must be created. For example:

Example 1.3:

"Imagine **you** are standing at the entrance to a park, to your **right** is an information centre, to your **right** is a rubbish bin."

"***You – Rubbish bin – information centre***"

However, as both the information centre and the rubbish bin are only referred to in relation to "you", we cannot be sure of the location of the rubbish bin and the information centre in relation to one another and can only surmise that they are to the right of "you". Therefore, it is possible that the model we create is:

Example 1.4:

a) "***You – information centre – Rubbish bin*** "

or

b) "***You –Rubbish bin – information centre***"

The lack of spatial information regarding the location of the rubbish bin in relation to the information centre allows for the creation of two possible models, both of which are viable and are consistent with the information given in the premises. Both models are equally plausible. It is important to note that the location of "you" remains the same in both models. Throughout the thesis we refer to a situation with more than one viable model as "ambiguous".

Note: *From here on we will refer to an object/subject that remains in the same position regardless of the location of the other objects/subjects and an "anchor" object/subject (the object/subject in the premises which is not in an ambiguous position) while the other objects/subjects may be in different positions depending on the model created by the reasoner. Additionally, we refer to these objects/subjects being in* *“ambiguous” positions within the array. The reason we refer to the underspecification as an ambiguity is that it has led to the creation of two equally plausible yet competing interpretations, as either model could be true but not both, which is a defining feature of ambiguity, be it lexical, syntactic or another type (e.g., Deemter & Peters, 1996; Stojnić et al., 2018).*

Determining the location of “you” within the two models is relatively simple as in any solution which holds true to the premises you are always to the left of both the information centre and the rubbish bin. If asked, “ Which object/subject” is furthest to the left?” the only correct answer would be “you”. However,the question “Which object/subject is furthest to the right?”, has two equally plausible answers. A reasoner, using calculus-like logic, may infer that there is a 50/50 chance of the information centre being furthest to the right despite there being no clear answer as both solutions share the same probability and are dependent on one another’s relation to the anchor. However, it is known that people appear to rely on a variety of heuristics in making judgements about such probabilities (Tversky & Kahneman, 1983)

- **Representativeness**: used when reasoners are asked to judge the likelihood of what class an object/subject belongs to.
- **Scenarios**: plausibility of a particular development (e.g., something moving from one place to another.)
- **Adjustment from an anchor**, employed in numerical prediction (usually only used when numerical values are available)

When reasoners are working to interpret the premises to form a mental model it may be the case that, if they reach the conclusion of there being a 50/50 chance of an object/subject being in one place or another, the reasoner uses an object/subject's representativeness (e.g., "Object A belongs to the class of Objects on the right") and an evaluation of the scenario (e.g.," Is it plausible that Object B is directly to the left of Object C"). We discuss the evaluation of the scenario and its influence on mental model formation in detail later. Adjustment from an anchor is a less useful heuristic for the creation of a mental model as, the anchor (the object/subject which is not in an ambiguous position) has a 100% change of being in it's described location, whereas the other object's subjects may vary in terms of change (e.g., if there are 3 objects/subjects described in the premises of an ambiguous situation, then the non anchor object/subjects have a 50/50 chance of being in one location or another).

Among the key properties of models is that their structure corresponds to the structure of what they represent (like a visual image), and thus that individual entities are represented just once in a model. Therefore, if there are two equally viable models, the reasoner would have to envision the alternative possibilities by two alternative models (Johnson-Laird, 1994;Ragni & Knauff, 2013). It is understood reasoners generally only construct a single model and overlook other potentially viable ones (Johnson-Laird, 1994; Rauh et al., 2005). It appears that reasoners prefer to construct one model and "check" it for errors. As "erroneous conclusions will tend to be consistent with the premises rather than inconsistent with them", as if there are multiple solutions the erroneous conclusion that only one solution is correct is still erroneous. A model which "passes" such "checks" becomes a reasoner's preferred template from which to make further inferences.

Yet, it is currently unknown what occurs if a reasoner's preferred model is probed in a way which is consistent with another viable model, but inconsistent with their preferred model. For example, probing a reasoner's understanding of the premises in *Example 1.3*, with a question such as " Is the information centre to the left or the right of the rubbish bin?" may result in a reasoner answering either "left" or "right" depending on which of the viable models (*Example 1.4; a or b*) they prefer. A further unknown is, when prompted to state if the premises do not give them enough information, if reasoners even consider the alternative, non-preferred model. The main question posed by this Thesis is whether or not reasoners can accurately determine if a model has more than one viable solution, if not why and if they can what is the process involved.

The construction of a mental model.

It is important to understand what might influence a reasoner's construction of a mental model. Mental model theory makes assumptions on how reasoners construct a model. One crucial assumption is that the reasoner creates an initial model made up of the premises, and if there is more than one viable model, generates alternative models sequentially (Johnson-Laird & Byrne 1991). Johnson-Laird and Byrne describe this process by breaking it down into three phases; comprehension, description, and validation. During model comprehension, a reasoner uses their knowledge of how the world works to construct their initial model which is then validated by checking the premises. However, mental model theory assumes that only this initial model is required, even if there are additional adjustments that need to be made, and that the premises may be forgotten (Mani & Johnson-Laird, 1982).

There are different theories as to how the construction of mental models occurs. For example, while not drastically different from Mental Model Theory, is that of Preferred Mental Model theory (Ragni & Knauff, 2013). It suggests different terminology for the model construction phases, these being; construction, inspection, and variation. As with Johson-Laird and Byrne's comprehension phase, the In the model inspection phase, a "generally true" version of the mental model is constructed using the premises. This initial model is then inspected by the reasoner to make preliminary conclusions and during this phase the reasoner is able to find relations not directly stated by the premises (Ragni &

Knauff, 2013). The inspection phase described by Ragni and Kauff is not drastically different from Johnson-Laird & Byrne's description phase. However, Johnson-Laird & Byrne's description phase assumes the reasoner's preliminary conclusion functions as a description of the model as a whole (Johnson-Laird & Byrne, 1991).

The last phase, model variation phase, called validation by Johnson-Laird & Byrne, involves the reasoner attempting to find an alternative but valid model, or a model in which the conclusion does not hold true with the premises. Assuming the reasoner finds a false model, they reject it and create another model. If they cannot find an alternative model in which the premises hold true, their initial model is valid and conclusions based on it must also be true. If at any point a contradiction is found, the reasoner returns to the first phase until all possible models are generated and tested (Johnson-Laird & Byrne, 1991).

Preferred model theory differs from the standard mental model theory, not only in the names of the phases but also in attempting to better explain how reasoners keep track of all the possible models. For example, Byrne and Johnson-Laird (1989) speculated that reasoners attempt to consider all models that can be created using the premises but that they fail to hold all of these models in their working memory. With relatively simple premises that result in more than one model (e.g., *Example 1.4*) holding multiple models in memory may not be hard but becomes incrementally more difficult the more objects/subjects that are added to the premises. An alternative to reasoners holding all potential models in working memory, and failing, is that models may be represented symbolically with mental "annotations" as to where an object/subject might be (Vandierendonck et al.,1999). Additionally, it hasalso been shown that models might be constructed through the use of partial models, excluding ambiguous elements, which are then revised (Vandierendonck et al., 2004).

It is also important to consider what influences the construction of mental models apart from the premises. Nejasmic et al. (2015), investigating how mental models are revised based on an object's physical properties, found that the model created by the reasoner is influenced by “asymmetry” between a reference object (e.g., “*you”* in *Example 1.2*) and located object (e.g., the “*Rubbish bin*” in *Example 1.2*). Further, by comparing “movable” and “immovable” objects, Nejasmic et al. (2015) found that some physical properties (e.g., large, heavy vs. small, light) of the objects used in premises can influence the mental model a reasoner constructs. However, it is not yet certain what precise properties influence model construction as it was found that “some but not all” properties influence model preference (Nejasmic et al., 2015). For example, it was found that object weight did not significantly influence the outcome of the preferred model but that the size of the object did and that this was further compounded by whether the object was movable or immovable. The prior research indicates that reasoners, when given new information (e.g., an additional premise or description thereof), revise their model and that this revision is influenced by assumptions about an object's physical properties (Nejasmic et al., 2015). However, what is not yet known is if this interaction of model revision and the physical properties of a mental object influence if a reasoner can determine if a model has more than one viable solution/order. Therefore, the empirical chapter’s of this thesis generally use an “an equal” approach to object/subject size, movability, etc. by using sailing ships (which share the same properties) but with different names.

The creation of a mental model based on spatial information is influenced by how objects in those models behave in the world (Nejasmic et al., 2015). However, this process appears to have other, potentially compounding, issues. The presentation order of the objects/subjects within the premises also influence a model’s construction (Ehrlich & Johnson-Laird, 1982), whereby continuous descriptions (e.g., “*A in relation to B, C in relation to B or A, D in relation to C*”) were easier to for reasoners to make inferences about than discontinuous premises (e.g., “*A in relation to B, D in relation to C, C in relation to B or A*”). Furthermore, it was demonstrated that reasoners have difficulty in general when the premises create indeterminate orders (Mani & Johnson-Laird, 1982). It is also important to note that reasoners make fewer errors when the premises describe more exact relations such as “next to” or “overlaps with” (non-transitive) rather than “left of ” or “in front of” (transitive), (Goodwin & Johnson-Laird, 2008; Knauff & Ragni, 2011). It is thought that non-transitive relations are more easily constructed into a viable model by the reasoner.

Another important element of the construction of a mental model occurs when the premises describe an ambiguous situation. As each model created represents one viable possibility, rather than one single logical outcome which is the case for logical inferences, the reasoner may have to draw more than one conclusion from the premises. The Conclusions which are based on premises and assumptions such as “information centres do not float in mid air”, allow for the creation of a model in which all those premises hold true. However, even if a model requires the reasoner to construct the model in such a way that “the information centre is floating in mid air”, the model remains viable but is simply “less preferred” (Rauh et al., 2005). The model created by a reasoner can only be rejected by a reasoner if they find a counterexample e.g., a situation where the premises are correct but the conclusion is not, or a conclusion which does not hold true to the premises (Schroyens et al., 2003). If a model requires one of the objects/subjects to act in a way which is contrary to real world rules e.g. A floating information centre or a rubbish bin occupying the same space as a person, then this model is less preferred.

It is also important to understand that premises which describe the location of objects/subjects in real world space are handled differently than the orders of objects/subjects in “non-space”. Many studies have shown that the mapping of non-spatial situations on spatial relations is deeply rooted in human cognition. For example, Gattis & Holyoak (1996) showed evidence of reasoners using intermediate mental representations, which enable them to move from fully visuospatial representations to more abstract mappings. This implies that when a reasoner is faced with a non-spatial situation for which they must construct a mental model, that this mental model is “built” using spatial, visual images. However, Knauff (2009, 2013) has argued against this view and proposed that visual images are not relevant for reasoning and might even impede the process. Instead, Preferred Mental Model Theory (Knauff et al., 2013) suggests that thought processes which are more abstract than pictorial images but more concrete than linguistic representations. Therefore it is likely that spatial and non-spatial situations are processed differently.

How mental models are constructed by reasoners have been studied extensively (e.g., Rips, 1994; Johnson-Laird, 2001; Rauh et al., 2005). Research into mental models typically aims to investigate the reasoning processes involved in creating a preferred solution for an order of objects/subjects, be they ambiguous or not, with a focus on understanding the reader's inference/reasoning processes. It is known that reasoners, in an underspecified situation such as our example, would end up preferring one interpretation over the others (Goodwin & Johnson-Laird, 2005; Nejasmic et al., 2015).

If the premises describe a model with multiple solutions it is known that the reasoner's preferred solution is accepted rather than the others deactivated (Rauh et al., 2005). Mental model research has also found that people use a variety of strategies for constructing and subsequent deduction of a mental model (Schaeken et al., 2000). Readers also have comprehension difficulty the more premises and entities contained in the premises (Goodwin & Johnson-Laird, 2005). Additionally, there are semantic and syntactic influences on how a preferred model is created (Rips, 1994; Johnson-Laird, 2001). It is known that people prefer to create a parsimonious order of objects/subjects while still obeying all the premises (Jahn et al ., 2007). In turn, the models that are preferred tend to be simpler and require fewer assumptions, as they are easier to understand and remember (Byrne & Johnson-Laird, 1989; Goodwin & Johnson-Laird, 2005). Importantly, people also prefer to create a coherent order, i.e., one which does not violate a person's prior knowledge of the world (Goodwin and Johnson-Laird, 2005) and they prefer to create an order that is useful for the situation (e.g., answering a comprehension question) (Nejasmic et al., 2015). For example, if a reader is presented with a stimulus such as the one in *Example 1.3*, and are asked about their mental model with a question such as “Which object is furthest left?” either of the two viable models are useful enough to make an inference from. Lastly, It is known that models are created “bit by bit” (Johnson-Laird & Byrne, 1991).

However, using this prior research does not allow us to ascertain precisely which mental model in Examples 1.3 & 1.4 would be preferred by a reader. All viable solutions do not violate the premises in Example 1.3 and all of these solutions are useful (Nejasmic et al., 2015) for making inferences about the position of the objects within the model. All we know is that a reader would prefer one of these interpretations over the other and use their preferred model to make inferences. Furthermore, we do not know if a reader of the situation described in Example 1.3 would be aware that multiple solutions exist.

Differences between this Thesis and prior mental model research

Typically, mental model research does not investigate whether or not a person is aware that the premises describe a situation that has multiple viable solutions. Mental model research focuses on which solution participants preferentially report (e.g., Rauh et al., 2005), while other methodologies (e.g., Jahn et al., 2007) focus on whether participants can create models in which all premises can hold true. In the case of our example a participant in Jahn et al.’s 2007 study could report either of the viable solutions we outlined as none of these contradict any of the premises. However, while it is known that reasoners generally do not consider non- preferred models (Johnson-Laird & Byrne, 1993), it is not known whether participants consider “non-preferred” models (Rauh et al., 2005) when prompted to detect them (e.g., if there is enough information for a definite answer). Participants may be unaware that the model they construct is not the only viable one. We therefore do not know if participants can detect ambiguity arising from described situations.

Additionally, mental models are investigated presenting the premises in short isolated sentences such as *"to your **left** is a rubbish bin*" (Nejasmic, Bucher & Knauff, 2015; Barkowsky, Freksa & Knauff, 2005; Knauff et al., 2004; Rauh et al., 2005). But these sentences do not reflect everyday reading and readers do not typically read isolated premises in a description of a situation. The empirical chapters in this thesis therefore use premises which have been embedded into larger, more "natural" paragraphs, examples of which can be seen in Chapter 2. These premises describe various situations, some of which have a clearly defined inference for participants to make, while others have multiple solutions for participants to detect ambiguity.

Furthermore, it is also often the case that participants have access to the premises during reasoning (e.g., Goodwin and Johnson-Laird, 2005), however, in this thesis we test participants' ability to recall premises from memory. While the memory recall aspect in the studies in this thesis are not new (e.g., Jahn et al., 2007; Rauh et al., 2005), little is known about how ambiguity detection from memory might work. We therefore test participants in a way which requires memory recall of the model and/or premises. If, as speculated by Byrne and Johnson-Laird (1989), reasoners attempt to remember all possible models but fail to do so, it stands to reason that simply stating if there was more than one viable solution may be easier for the reasoner than constructing multiple (viable) models. Even if, as speculated by Ragni & Knauff (2013), reasoners have a "blindness for multiple models, meaning that people are almost blind to the existence of alternative models and basically treat multiple-model problems as though there were a single possible model", ambiguity detection from memory would be similarly difficult. This thesis therefore also investigates how ambiguity is detected from memory, but does not compare this to how it is detected when the reasoner has access to the premises. Furthermore, the encoding of premises, i.e., the reading of the premises, is also not studied in mental models. We therefore only probe participants' knowledge of the situations without them having access to the premises, which allows us to study the way in which participants encode the premises.

Also understudied is the content of the text. It is currently not known if the spatial situation described in Example 1 is processed differently from a non-spatial one. While studies investigating relations other than spatial ones, such as temporal relations have found that readers still prefer to create a single model (Goodwin & Johnson-Laird, 2005), non-spatial situations are rarely compared to spatial ones. As such it is currently unknown if non-spatial situations such as:

Example 2.0:

"*Toby is faster than Jacob. Sarah is slower than Toby"*

are processed differently from spatial situations such as :

Example 2.1:

"The horse was to the left of the cliff, the boat was to the right of the horse"

It is uncertain if spatial and non-spatial situations differ in terms of a reasoner's model comprehension and if ambiguity is detected at different rates.

In our exploration of how and when ambiguity is detected in different types of mental models further considerations have to be made on the readers themselves. If spatial models are more prone to semantic interference/guidance for example, then ambiguity may be harder to detect from spatial mental models compared to non-spatial ones. Semantic interpretation is often influenced by cultural biases (Djiwandono, 2006) whereby prior knowledge of the physical world and social interactions, based on the reader's culture, influences the interpretation of the text. Therefore, it is possible that culture plays a key role in the construction of preferred mental models when there are multiple, equally likely interpretations. It is also important to note that we assume that the premises are assumed to be "reliable" by readers. As it has been shown that there are significant differences between reliable and unreliable information sources in the construction and preference of mental models (Wolf, et.al., 2012).

While this thesis does not directly address potential cultural biases, and accounts for them through pre-screening there are additional considerations to be made on the formation of spatial models themselves. In terms of spatial situations and semantics it is important to note that writing and reading in a certain direction is able to induce how a spatial model is interpreted (Román et al., 2015) and can change depending on whether or not the participant is a monolingual or bilingual person (Román et al., 2013). Román et al., 2013 investigated this using Spanish and Moroccan participants which were presented with audio stimuli and were asked to draw the spatial premise the stimuli described. Spanish participants showed a preference for drawing the lamp on the left and the TV on the right after hearing the stimulus “The table is between the lamp and the TV”, whereas Moroccan participants, who read and write in Arabic (right to left), tended to draw the lamp on the right and the TV on the left. Therefore it would be safe to assume that a preferred mental model could change depending on the reading direction of the participants native language and in which language the experiment is in. Similar preferences found in spatial orientation have also been found in native writers of Hebrew (Andrews et al., 2013). Therefore it is known that there are semantic influences on how a situation may be interpreted by reasoner but in order to mitigate potential cultural/linguistic influences on model preference (Andrews et al., 2013; Román et al., 2015; Román et al., 2013) we only tested native english speakers residing in the UK.

While it is hitherto unknown if spatial situations are processed differently from non-spatial ones, situations which describe non-spatial situations may also be influenced by a participant’s native language. Social biases can also be found in (Maass et al., 2007), the participants were asked to rate a football goal in terms of strength, speed and beauty. It was found that Italian participants showed a proclivity to interpret an athletic performance (the scoring of a goal) as stronger, faster, and more beautiful if presented with a left-to-right rather than right-to-left trajectory. While this an instance of spatial preferences influencing non-spatial descriptions and not directly studied in this thesis, it is still noteworthy that social interpretations of an event are subject to cultural biases.

Ageing

Mental models and preferred mental models have been studied extensively, but mostly in young adults. Working memory declines as we age (Bopp & Verhaeghen, 2007; Borella et al., 2008; De Beni et al., 2004; Fiore et al., 2012 ; Park et al., 2002). It is known that visuo-spatial working memory is inherently linked to spatial reasoning tasks (Gyselinck et al., 2009; Gyselinck et al., 2007; Wang et al., 2018) and that mental models are “held” in working memory (Baddeley, 1998). A decline in working memory capacity may therefore influence the ability of a person to construct and reason with a mental model. Furthermore, it has been found that older people struggle to integrate information into a mental model (Copeland & Radvansky, 2007) and that older readers are worse at inhibiting irrelevant information than younger readers (Hasher & Zacks, 1988; Hamm & Hasher, 1992).

However, age-related declines in working memory or model integration ability may not influence the creation of a mental model. For example, competence in making inferences from text appears to be no different between younger and older people (e.g., Light, 1988; Radvansky et al.,1990). Additionally, white older people struggle to retrieve correct and relevant information from memory (Gerard et al., 1991) this does not interfere with the ability of older people to create a mental model (Radvansky et al., 1990). Furthermore, once a mental model is created it can be retrieved by both older and younger readers without much difficulty (Gilbert et al., 2004). While declines in working memory capacity and subsequent difficulty in memory retrieval of premises cause some problems for older people, their ability to form and reason with a mental model remains intact. It has been found that older readers take significantly longer than younger readers to encode spatial model information, but that once a mental model is formed older readers are able to make inferences from the model as well as younger people (Kemper et al., 1993; Liu et al., 2017) . Lastly, older people often outperform their younger counterparts when it comes to rotating mental images (Dror et al.,

2005) and so may be better at dealing with a spatial mental model. However,it is currently unknown how age-related changes in memory may influence the ability of a person to detect ambiguity.

As we age, a decline in working memory causes readers to break down a text into smaller, more manageable, chunks (Swets et al., 2007). Therefore, an age-related decline in working memory may influence the processing of situational ambiguity by making it harder for the two premises to be combined efficiently (Copeland & Radvansky, 2007). This may result in a reader with a lower working memory capacity (or under a higher working memory load) to incorrectly recall the premises in a situation and create a mental model based on incorrect information. Alternatively, difficulty in memory retrieval (Hamm & Hasher, 1992; Hasher & Zacks, 1988) may cause older people to not consider more than one viable option, thereby not detecting ambiguity. However,whether this is the case is not yet known as ambiguity detection has not yet been studied across the life span.

However, despite ageing and ambiguity detection being unstudied, there is evidence from discourse processing to suggest that processing strategies may be different in older and younger readers. For example, as the premises used as part of the stimulus in this thesis are separated by irrelevant (to the creation of a mental model) text. It could be argued that the premises act like long distance dependencies, the processing of which is known to take longer in people with lower memory capacities (Nicenboim et al., 2015) and the general processing of text also slows with age (Myerson et al., 1990). Given that the ability of an older person to create a mental model does not decline (Radvansky et al., 1990), it may be that the process by which a mental model is created and comprehended involves different strategies in older people compared to younger people. However, it is unknown if different strategies are used, what these strategies may be, or if they are more or less beneficial for ambiguity detection.

The decline in working memory as people age (Chen, et.al., 2003; Fabiani, 2012 ; Klencklen, et.al., 2017) is known to reduce the ability to disambiguate lexical and syntactic ambiguity effectively (Yoo & Dickey, 2017; Christianson et al., 2006). Lower working memory capacity also directly influences a reader's ability to disambiguate syntactic ambiguities (MacDonald et al., 1992). This may indicate that, when reasoning with mental models, older people may experience additional difficulty when attempting to detect ambiguity. This further supports the notion that older people face difficulty when combining premises for the construction of mental models (Copeland & Radvansky, 2007). Declining working memory is also known to cause declines in text comprehension (Van der Linden et al., 1999;Schurer et al., 2020), especially when presented with alternative interpretations (Uekermann et al., 2008) which may make it harder for older readers to correctly construct a mental model of a situation described in text. However, whether findings from lexical or syntactic ambiguity can be applied to situational ambiguity is not known, although they serve to highlight potential processing differences as a result of working memory decline. It should be noted that declines in working memory is not limited to older people and also hinders younger people in their ability to retrieve relevant information (Slattery et al., 2013; Malyutina & den Ouden, 2016; Salhi & Bergström, 2020).

This thesis

Work on ambiguity resolution from text as well as how and why certain models are preferred over others have been studied extensively. However, there remain several gaps in knowledge as to how people realise that a situation is ambiguous. This thesis develops a novel approach to ambiguity detection. Firstly, while it is known that reasoners generally prefer to create a single model from which to make inferences (Goodwin & Johnson-Laird, 2005; Nejasmic et al., 2015), which is preferred over others rather than others "deactivated" (Rauh et al., 2005), it is unknown if reasoners realise that other viable options may be available if prompted. Therefore, throughout this thesis we test whether or not participants are able to detect ambiguity by prompting them to answer “there is not enough information” if a situation is ambiguous. Furthermore, we compare reader’s ability to detect ambiguity and their ability to make inferences from unambiguous situations.

Secondly, it is not known if spatial situations are interpreted differently from non-spatial ones as these are typically not compared to one another directly. We therefore compare participant’s ability to detect ambiguity and make inferences about unambiguous situations between situations using spatial and non-spatial models. It is possible that spatial and non-spatial situations differ in their construction due to semantic interferences (Knauff & Johnson-Laird 2002; Knauff & May, 2005). However, as the direct comparison between the two has not yet been made in prior literature this element is somewhat exploratory. The difference between spatial and non-spatial situations is addressed in every chapter of this thesis.

Thirdly, we aimed to investigate how ageing influences inferences made about unambiguous mental models and how ageing may influence the detection of ambiguity. As working memory resources are likely lower in older people (Bopp & Verhaeghen, 2007; Borella et al., 2008), having difficulty in memory retrieval (Hasher & Zacks, 1988; Hamm & Hasher, 1992) and difficulty in mental model generation (Copeland & Radvansky, 2007) compared to younger people, it may be the case that older people use different encoding and reasoning strategies (Swets et al., 2007) or compensatory mechanisms (Piefke et al., 2012). We therefore investigate not only the ability of older people to detect ambiguity and make inferences from mental models but also the reading behaviour (Chapter 4 and 6) of older people in order to glean information on the encoding of premises and whether eye-movements differ when encoding ambiguous premises compared to non-ambiguous ones. Furthermore, we use different question probes in order to investigate potential benefits to mental model processing and ambiguity detection to both older and younger people (Chapter 6).

Lastly, in Chapters 4 and 6 we explore the encoding behaviours of participants when reading situations, be they ambiguous or not. We further delve into these findings in Chapter 6, which is far more exploratory than the previous chapters, as we investigate if encoding and potentially reasoning strategies differ depending on how a participant's mental model is probed. We also investigate if different strategies may be used to better detect ambiguity.

Chapter 2: Exploring Ambiguity in Situational Models: Evidence from Reading Comprehension

In keeping with open science, the full reproducible code, analysis, pre-processing, and data are available via the public OSF page which can be found here.

https://osf.io/a3w2y/?view_only=e8867c665a144eddb5f7c3ca6745c1f9

Abstract:

Ambiguity in written language has been studied extensively, however, the concept of situational ambiguity (ambiguity that arises from underspecification in the situations described in the text) remains under-studied. Previous research on mental models (e.g., Goodwin & Johnson-Laird, 2005) has focused on which interpretation of the ambiguity participants prefer rather than on if/when readers become aware of situational ambiguities and how these ambiguities are then resolved. We present a novel naturalistic reading methodology for studying readers' ability to "detect" if an object/subject order was ambiguous or not. This study presents a methodology for investigating ambiguity "detection" using ambiguous and unambiguous subject/object relationships embedded into "natural reading" paragraphs. The object/subject orders were similar to mental models research (Goodwin & Johnson-Laird, 2005; Nejasmic, Bucher & Knauff, 2015). Questions were presented after the text, probing participants' understanding of the texts' situation model. We also investigated if the type described in the text (spatial vs. non-spatial) matters. We present evidence that situational ambiguity often goes unnoticed by readers. Furthermore, situational ambiguity becomes apparent to readers when probing their understanding of the described situation and their created situational model more often with non-spatial situations than spatial situations. We suggest that readers hold two independently retained situational relationships from the text which are combined during questioning which leads to a higher question reading time for ambiguous models. Furthermore, we find that spatial situations are harder to reason about than non-spatial ones regardless of their ambiguity.

1.0 Introduction

Premises describing how objects/subjects relate to one another and the resulting mental models have been studied extensively (e.g., Rips, 1994; Johnson-Laird, 2001; Rauh et al., 2005). Research into mental models aims to investigate the reasoning processes involved in creating a preferred solution for an order of objects, be they ambiguous or not, with a focus on understanding the reader's inference/reasoning processes. To do so, mental model research often uses globally ambiguous stimuli that are consistent with multiple solutions to study which solutions reasonsers prefer (e.g., Goodwin & Johnson-Laird, 2005). For example, if presented with the situation:

Example 1:

- The coffee cup is to the left of the phone.

- The book is to the right of the coffee cup.

a participant would be asked to report the order in which the items are organised. In this example, there are two possible solutions [coffee cup, book, phone; coffee cup, phone, book].

In existing mental model research participants are not required to report if the description is ambiguous. Mental model research instead focuses on which interpretations (or mental models) of ambiguous descriptions participants preferentially report (e.g., Rauh et al., 2005). Other methodological paradigms (e.g., Jahn et al., 2007) focus instead on whether participants can create models in which all premises can hold. As such it is not known whether participants consider “non-preferred” models (Rauh et al., 2005) when deciding if a text creates multiple viable solutions. Participants may be unaware that the model they construct is based on incomplete/under-specified information if unprompted to reason with multiple models and need only state if multiple models exist. Furthermore, mental models are usually investigated by presenting the premises in short isolated sentences e.g., *"The coffee cup is to the left of the phone and The book is to the right of the coffee cup"* (Nejasmic et al., 2015; Barkowsky et al., 2005; Knauff et al., 2004; Rauh et al., 2005). In everyday reading, however, readers will not typically be presented with isolated premises when a situation is described. Instead, premises are embedded in continuous text. To simulate a more naturalistic reading situation, premises will be embedded in paragraphs in the current study (see Example 2 and Figure 5). These paragraphs contain additional information that is related to the described situation but not relevant to the actual reasoning task.

Example 2:

"The table was laden with many fanciful foods. The servants had prepared a great feast in celebration of the harvest moon. <u>The apples were to the right of the pears</u>. The walls had been decorated and the chairs were built from old threshing boards. <u>On the table, the oranges were to the left of the apples</u>. It was certainly going to be the grandest feast since the coronation of the king."

Despite the change in tense and wording compared to a premise from Example 1, the situation described by the two premises remains ambiguous. We currently know little about how situational ambiguities are processed when presented in such more "naturalistic" reading tasks. In this study we therefore presented readers with two paragraphs of text, each describing a situation containing two premises, similar to the one shown above in Example 2.

Mental models are also typically studied in a way that allows participants to access the premises during reasoning (e.g., Goodwin and Johnson-Laird, 2005) or tests a participant's knowledge of the model from memory (e.g "*Is there any layout for which all the assertions are true?"* (Jahn et al., 2007), or "inserting" an additional premise after the first two have been presented (e.g., Rauh et al., 2005). However, model comprehension is rarely investigated from memory (e.g., "Are the apples directly to the left of the oranges?") and less still is known about how ambiguity detection from memory might work. In this study, we investigate if readers can “detect” ambiguity if they no longer have the text available to them and must rely on memory of the described situation.

It is important also to consider how the “style of the text may influence a reader’s interpretation of the text and the model/s they create from the given information. While at first, it could seem as though situational ambiguity is not influenced by semantics, some situations may lead readers to prefer one interpretation over another or deactivate a “less appropriate” interpretation. For example an ambiguous situation such as:

Example 4:

“The knife is to the left of the spoon”

and

“The spoon is right of the fork”

Here there are two viable solutions; “fork, knife, spoon” and “knife, fork spoon”. If asked “which object is furthest right” a reasoner with knowledge of table etiquette may prefer the solution “fork, knife, spoon”. In this example a reasoner’s preferred solution is accepted rather than others deactivated, which is known from prior work in mental models (Rauh et al., 2005). However, a reasoner’s preference to construct the “fork, knife, spoon” model could cause the reasoner to neglect the “knife, fork, spoon” model and as a result not detect that the described situation is ambiguous. By contrast, if we consider a model such as:

Example 5:

“The horse was to the left of the cliff, the boat was to the right of the horse”

We again have two viable models; “*horse, cliff, the boat*” and “*horse, boat, cliff*”.

Here context could inform a reasoner that Boats and Cliffs are both objects found near water, while horses are land animals (Matthew, 1926). As such the “horse, boat, cliff” model may be less preferred, as it could imply that the horse is in the water or the boat is on land which does not fit into a “normal” worldview and/or assumes that the horse or the boat are outside of their natural setting which may not be a preferred model (Goodwin and Johnson-Laird, 2005). This could result in the deactivation of one viable model due to semantics or context, leading the reasoner to not consider it viable ergo not detecting ambiguity in the situation.

This semantic interference can impede the process by which a reasoner creates a mental model (Knauff & Johnson-Laird 2002; Knauff & May, 2006). However,additional problems may arise from a situation where there are far fewer semantic clues. For example:

Example 6:

“Toby is faster than Jacob. Sarah is slower than Toby”

The reader may be unaware that it is not possible to tell who is the slowest and simply choose one of the two available solutions; “***Toby**, **Sarah, Jacob**” or “**Toby**, **Jacob, Sarah”*** (if ordered *with the fastest (Toby) on the left and slowest (Sarah or Jacob on the right).* In both instances, either solution is acceptable (or “good enough”) but readers may be unaware of the alternative.

There is no information that could sway a reader from one interpretation to the other even though the example presents a great deal of information. Both solutions, “*Toby, **Sarah, Jacob**” or “Toby, **Jacob, Sarah”,*** are acceptable. It is not known if by preferring and applying one solution over another (Rauh et al., 2005), or ignoring one in favour of another, ambiguity can still be detected. To test if participants “detect” ambiguity they have to state whether “Sarah is fastest”, “Toby is fastest” or whether “there is not enough information” thereby acknowledging that both solutions are possible but it is not possible to determine one solution over others. It is currently unknown if participants can detect ambiguity in this way.

In this exploratory study, we investigated if readers can “detect” ambiguity in situational models that were embedded in text.

Furthermore, it is unknown if ambiguity is more easily detected in situations described through objects/subjects with fewer “real world” biases, e.g. “Sarah, Jacob, and Toby vs. knife, fork, spoon”. It is possible that clues, such as table etiquette in Example 4, “steer” the reader to create a preferred model over another. However, personal biases might influence the interpretation of non-spatial situations(Andrews et al., 2013; Maass et al., 2007) but to a lesser extent. Consider Example 6, where there are two viable speed orders of ***Toby***, ***Sarah,*** and ***Jacob***. A singular reasoner might know a very fast “***Jacob***” and might construct a model in which “***Jacob”*** is faster than “***Sarah”*** but there are no clues that would guide every reader to construct a model in which “***Jacob”*** is faster than “***Sarah”***. Semantically a “***Toby***” is no faster than a “***Jacob***” or a “***Sarah***”. It is currently unknown if a “non-spatial” situation (Example 6) is any easier to construct and reason with than a spatial one (Example 5), nor is it known if ambiguity is harder/easier for participants to detect in one or the other type of situation. The current study also addresses this by comparing how readers deal with the type of situation, spatial situations (Examples 4 & 5), and non-spatial situations (Example 6).

Because of these possible influences on model preference in spatial and non-spatial situations, the stimuli in this study were written in a way to limit their influences. For example, in a spatial situation describing three objects, none of the objects are commonly found on walls, e.g. paintings, as the wall may create a barrier that prohibits an object from being placed beyond it. Only if all objects are placed on a wall e.g. a clock, a painting, and a calendar would such a situation be used as a stimulus.

It should be noted that other factors may play a role in determining model preference. For example, time is represented in space as left (further back in time) and right (forward in time). Ulrich and Maienborn (2010) found that participants responded faster when relating backward in time with a left response and forward in time with a rightward response than when the response mapping was reversed. This suggests that semantics based outside of the text may influence the creation of a model. If we return to Example 4 and the knife, fork, and spoon, left and right-handed people may arrange the model differently based on their hand dominance, or that native language reading/writing direction influences how a model is constructed (e.g., Castelain & van der Henst, 2021; Lopiccolo & Chang, 2021). Additionally, a spatial model derived from text is not always fully constructed to decrease cognitive load (Lin & Matsumi, 2022).

While these findings help us to understand why a person may choose to create one model over the others, it is currently not known if readers can accurately detect whether the situation is ambiguous. The main aim of this study is to establish the use of premises describing a situation, embedded in a body of text rather than on their own, and subsequent probe of a reader's situation model, as a viable method to research if ambiguity is detected and processed rather than which end model is created. Participants in mental models research are often aware that multiple solutions to a given model may exist as they are either prompted (Rauh et al., 2005) or are asked which model they prefer (e.g., Goodwin & Johnson-Laird, 2005). Prior research also focuses on which model participants constructed first (e.g., Jahn et al., 2007) by testing participants' ability to create models in which "all premises can hold". Although, in the latter study (Jahn et al., 2007) participants were not explicitly made aware that the premises presented created an ambiguous order and did not have the option to state that there are "multiple models", making it unclear if participants were aware of the ambiguity. The task presented in this study focuses on whether or not ambiguity is detected when appropriate.

1.1 Differences between mental model research and secondary research questions

Earlier research into mental models investigated how/why reasoners create preferred models from ambiguous orders of objects/subjects described in short premises (Nejasmic, Bucher & Knauff, 2015; Barkowsky, Freksa & Knauff, 2005; Knauff et al., 2004; Rauh et al., 2005). However, It is currently unknown whether or not such ambiguity can be detected by readers at all. It is important to note that there are several differences when comparing the task presented in this study and prior research on mental models.

The task presented in this paper made use of two paragraphs, each containing a body of text unrelated to the task but also containing the premises. This was done to better “simulate” natural reading. The paragraphs containing the premises were written such that the two paragraphs presented to the participants could be displayed in either order and followed a common “theme” (e.g., describing a boat race in terms of the position of the boats and the popularity of the boat race compared to other events).

The task required readers to memorise the premises embedded within the text, i.e. the text describing the situation was not available during their reasoning process. While some earlier studies did involve a memory aspect (e.g., Jahn et al., 2007), they did not investigate the ability to detect ambiguity from memory. Given that people create mental models best suited for the completion of a task (Nejasmic et al., 2015), it may be the case that a complete model is not created until a question probes ambiguity (or the location of an object/subject in the unambiguous model). Our approach requires participants to reason from memory allowing us to conclude the natural processing, disambiguation, and interpretation of ambiguous texts. A systematic manipulation of memory requirement, however, is beyond the scope of the current project.

The study made use of different types of situations, describing objects in space (e.g., X left of Y, Z right of Y) as well as situations describing non-spatial situations (e.g., X richer than Y, Z poorer than X). It is currently not known whether one type (spatial or non-spatial) of a situation is easier to interpret than another(E.g. non-spatial easier than spatial), or if ambiguity is easier to detect in one or the other.

While mental model research has investigated relations other than spatial ones, e.g., temporal (Goodwin & Johnson-Laird, 2005), it has been found that readers create a single preferred model. Therefore it may be the case that situations described spatially are interpreted no differently than non-spatial situations. However, the construction of spatial and non-spatial models may be influenced by different semantic cues (Rauh et al., 2005), in turn, these may lead to one model being preferred over others which are not considered by the reasoner. It is currently not known if semantic interference plays a role in model creation and as a consequence ambiguity detection.

Our ambiguity manipulation creates two versions of the same paragraph, one with two potential models (ambiguous) and one with a single model (unambiguous). If readers create a single, preferred model from which to reason (Knauff & Johnson-Laird 2002; Goodwin & Johnson-Laird, 2005; Knauff & May, 2006), then it is likely that ambiguity goes undetected in most cases as the reader makes inferences from the viable model of their making rather than realising that it was not the only possible solution. However, there should be no differences in processing time for ambiguous and unambiguous stimuli as both contain the same number of premises and objects subjects that are described (Goodwin & Johnson-Laird, 2005) but how this translates into reading/reasoning time when the text is no longer available is not yet known.

2.0 Method

2.1 Participants

We recruited 180 (Mean age= 28 years; SD= 10.3 years 116 Female, 1 other, 63 Male) native English speakers via Testable-Minds (https://minds.testable.org/) (100) and undergraduate students at Bournemouth University (80). Participants were compensated $2 or 0.5 SONA research credits for participating in the study**.** Participants were naïve as to the ambiguity manipulation and were not informed of any underspecification in the text before the start of the experiment. However, participants were prompted to answer "there is not enough information" if there was more than one viable model. Participants were informed that they would be reading descriptions of "social"(non-spatial) and spatial situations. The study was approved by the Bournemouth University Research Ethics Committee (ID 27563). Each participant was informed of the experimental procedure and provided consent. Due to potential cultural biases (Andrews et al., 2013; Maass et al., 2007) in interpreting the texts, the study was limited to native English speakers living in the UK.

2.2 Materials and Design

The experiment had a 2 (paragraph type: non-spatial vs spatial) x 2 (ambiguity: ambiguous vs unambiguous) within-item design. Each text consisted of one non-spatial paragraph and one spatial paragraph. Paragraph order was counterbalanced.

The stimuli consisted of 24 pairs of paragraphs (items), in 4 conditions (ambiguous vs unambiguous) & (Spatial vs non-spatial) in each pair, one paragraph described a spatial situation, while the other described a non-spatial one (see Figure 5 for an example). The paragraphs within each item were consistent with the item's theme (for example a spatial description of the location of art in a gallery and a non-spatial description of monetary donations to the gallery) and were written such that they could be read in either order. Participants were presented with the paragraph pair, the order of paragraphs within each item was counterbalanced (e.g ambiguous spatial paragraph first, unambiguous non-spatial paragraph second) and each of the 24 items was presented in pseudorandom order to the participants (e.g Trial 1, item 22. Trial 2, item 3). The spatial and non-spatial paragraphs within an item (e.g see Figure 5a), were always displayed together and only differed in terms of order (spatial first, non-spatial second; non-spatial first, spatial second), ambiguity (e.g ambiguous spatial first, unambiguous non-spatial second; unambiguous spatial first, unambiguous non-spatial second). Every item was only shown once.

A series of three questions followed the stimuli once the participant had indicated that they had finished reading by pressing the space key. A single question would be presented along with three options (two viable model options, one correct for unambiguous stimuli, and “there is not enough information” correct for ambiguous stimuli). The order of the questions was determined by the order of paragraphs, e.g., “spatial first, non-spatial second (see Figure 5b for an example)” and followed by a third general comprehension question

Figure 5a.

An example stimulus: Item 3 in the configuration Unambiguous non-spatial first,

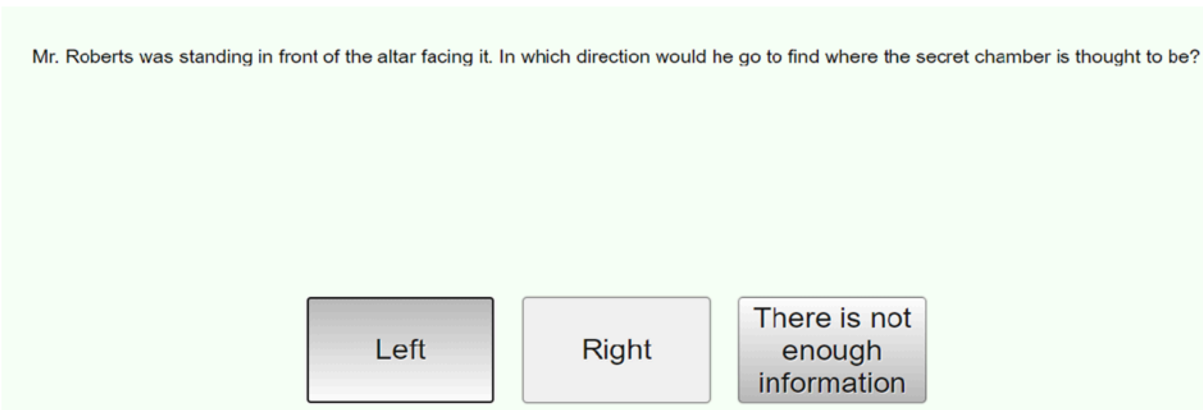
Benjamin the mechanic had only just arrived on the island. He had been called there on the behest of his friend Allan, who was also a mechanic, to fix the island's lighthouse. Benjamin was less skilled than Allan. After a long day of trying to repair the lighthouse, there was still much work to do to get it working. Allan decided to call Susan for help. Susan was more skilled than Allan. Surely the three of them would complete the work faster than just the two of them.

Spice Island lay just three miles offshore. There was a small weather station to the left the small dock used to load and unload supplies. The island didn't have a lot to offer in terms of comfort but there were a handful of people that called it home. There was also a supply store to the right of the docks. It was overpriced because everything had to be imported from the mainland. Many islanders didn't mind spending more on supplies if it meant avoiding a long trip to the mainland.

Unambiguous spatial second.

Figure 5b.

An example question for spatial Item 1, participants would select their answer using one of the answer buttons shown.



There were 24 stimuli items in 4 conditions. Therefore, to achieve sufficient power, 180 participants would be needed to arrive at the desired 1600 observations per condition (Brysbaert & Stevens, 2018). This ensures each item is observed at least 20 times in each condition. Results from a post-hoc power analysis, reported in the Appendix, indicated sufficient power for analysis of ambiguity detection.

The ambiguous paragraphs were constructed to be almost identical to the non-ambiguous paragraphs except that one premise was altered to make the situation ambiguous by changing the second premise. For the spatial premises, this would be done by changing one relationship to its antonym e.g. right to left. However, an antonym exchange wasn't always possible for the non-spatial paragraphs without compromising the grammar or general narrative of the paragraph. As a result, 4 non-spatial paragraphs differed from their unambiguous counterpart by more than 10 words while the rest differed by less than 10.

The paragraphs used all possible 3 object/person relationships equally, of the 8 ways a 3-object situation can be arranged 4 are ambiguous and 4 are non-ambiguous which are represented in the stimuli three times per relationship. The premises were un-nested (presented in two separate sentences) and presented in separate sentences within the paragraphs (see items Chapter 2 Appendix).

2.3 Stimuli Details

Readability analysis was done with the *readability* package for R with the Flesh-Kincaid measure (Kincaid, 1975) using the quanteda package (Benoit K, et.al 2018), while the word frequency analysis used the SUBTLEX-UK database (Keuleers, Lacey, Rastle and Brysbaert, 2011). The readability statistics are shown in Table 1. The stimuli were written for this project. The stimuli were not significantly different from one another in terms of readability and word count (see Table 1). The settings and named locations in the stimuli were taken from various stories (e.g Lovecraft, 2004 [1923]) and TV shows (e.g Wheadon, 1997) due to the names (e.g "Northport") sounding familiar to both citizens of the United

Kingdom and the United States of America (for future usage). However, it should be noted that the narrative in the stimuli bears little or no resemblance to the works from which the settings are taken.

Table 1:

Means table showing readability metrics

Condition	**Ambiguity**	**Word count**	**Word Length in letters**	**Readability**	
				Flesch-Kincaid	**Word Frequency**
			Stimuli		
Spatial	Yes	98.61 (6.41)	4.52 (0.24)	8.05 (1.42)	6.05 (0.13)
Spatial	No	98.96 (5.9)	4.56 (0.24)	7.99(1.45)	6.04 (0.13)
non-Spatial	Yes	94.07 (6.41)	4.52 (0.24)	8.18 (1.39)	6.06 (0.14
non-spatial	No	93.42 (8.31)	4.56 (0.24)	8.22 (1.41)	6.04 (0.13)
			Questions		
Spatial	Same	22.22 (5.65)	4.17 (2.17)	7.34 (2.28)	6.15 (1.17)
non-Spatial	Same	18.82 (4.34)	4.38 (2.25)	7.79 (1.71)	6.10 (1.23)

2.4 Procedure

Participants indicated their age, gender, any diagnosed disability that makes reading difficult, if they had normal, or corrected to normal vision, and if they were native English speakers. Then they were provided details of study participation needed for providing informed consent. Participants were informed that if there was more than one valid interpretation to answer with “there is not enough information”.

Participants read two practice items (4 paragraphs) at the start of the experiment showing ambiguous and unambiguous stimuli (presented first then second) and spatial and non-spatial (presented first then second).

Stimuli were displayed in Helvetica font (15.5 point) using Testable-Mind’s full windowed mode–tablets and phones were not allowed. The size of the text was calibrated using a 35.5cm by 40.6cm screen and always presented centred on the screen to preserve text spacing in all screen sizes as well as to keep the order paragraphs one above the other in the correct order.

The answer choices to the comprehension questions appeared below the text of the question on 3 buttons. The spatial questions were worded to be consistent with the perspective given in the stimulus (see Figure 5b.).

All 24 items were presented in random order. Participants completed the study in a single session. The experiment was designed to last between 25-45 minutes on average and participants could take short breaks when needed.

2.5 Analysis

We analysed 3 dependent measures: Stimulus reading time (SRT), question reading time (QRT), and answer accuracy with (Generalised) Linear Mixed Models using the lme4 package v.1.1-21 (Bates et al., 2015) in R v. 4.0.4 (R Core Team, 2021). We used contrast coding for Spatial Ambiguity (Ambiguous 1, Unambiguous -1), and non-spatial Ambiguity (Ambiguous 1, Unambiguous -1) in the analysis of SRT. For the question data dummy contrasts were used for Ambiguity (Ambiguous 1, Unambiguous -1) and Topic (non-spatial 1, Spatial -1). Therefore, the intercept indicates the grand mean of all conditions and the slope estimate indicates the difference from the grand mean. Both SRTs and QRTs were log-transformed. Random slopes for participants and items were included unless models failed to converge when they were included. The full reproducible code is available via the OSF (https://osf.io/a3w2y/?view_only=e8867c665a144eddb5f7c3ca6745c1f9).

3.0 Results:

We excluded 30 of 180 participants who appeared to simply respond as quickly as possible to finish the study. These participants had trial reading times of less than 1 second on more than 18% of their trials. Five additional participants were removed for having very long stimulus reading times (greater than 4 minutes) on over 18% of their trials. Trial data for the remaining 145 participants were excluded if they were read in less than 1 second (4.9%), indicating a misclick, or greater than 4 minutes (5.6%). Question trials with a response time below 1000 ms were also excluded (0.3%). Finally, 3 questions with RTs over 4 minutes were removed. During analysis, it was uncovered that two questions (one non-spatial and one spatial), from different items, were written such that participants got close to 100% correct in one case and close to 0% correct in the other. Therefore, the two stimuli and their corresponding questions were excluded from further analysis. With these elimination protocols, 72% of the data remained for analysis. The mean accuracy for the general comprehension question was 84.9% (SD= 14.3%, Range= 70%-100%), indicating that participants were reading for general comprehension as instructed.

3.1 Reading times of stimuli (Described situation)

Means and standard errors for SRTs are presented in Table 2.0 We found that there are no significant differences in the SRT in any of the conditions as we can see in Table 2.1

Table 2.0:

Mean stimulus reading times in ms, mean total QRT, and mean comprehension question reading time(SD in parenthesis).

Stimulus Condition	**SRT (ms)**[1]	**QRT (ms)**		
		Spatial	**non-Spatial**	**General**
Both unambiguous	61821 (33402)	9221 (9819)	9524 (9699)	5160 (5613)
non-Spatially unambiguous Spatially ambiguous	62460 (32183)	8908 (9896)	9172 (10482)	4760 (4720)
non-Spatially ambiguous Spatially unambiguous	61181 (34328)	8832 (8791)	8843 (8311)	4840 (5265)
Both ambiguous	61865 (33083)	9184 (10019)	9180 (10309)	4770 (4833)

[1] Table 2 shows means only, however we investigated stimulus reading time and found only null effects. These are available to view as part of the analysis code on the OSF which can be accessed via the link on the cover page.

Table 2.1:

LMM Results for SRT log(ms) as a function of the presence of which paragraph (non-spatial or spatial) was ambiguous or not or both.

Fixed effects	**log(SRT)**				
	b	**SE**	**df**	**t value**	**Pr(>|t|)**
Intercept	-0.0226	0.0182	2933	-1.245	.213
Non-spatial Ambiguous	0.0162	0.0182	293	0.89	.374
Spatial Ambiguous	0.0004	0.0258	293	0.002	.999
Both ambiguous	-0.0226	0.01822	293	-1.245	.213

3.2 Question reading time

We investigated QRT to assess whether or not the type (spatial or non-spatial) or ambiguity played a role in how long it took participants to answer the comprehension questions. We found a main effect of stimulus type whereby questions about spatial paragraphs took significantly longer to respond to than questions about non-spatial paragraphs (see Table 3), indicating that spatial situations are harder for participants to create models for. Additionally, there was a main effect of ambiguity with questions about ambiguous paragraphs taking longer to answer than questions about unambiguous paragraphs. However, this main effect is qualified by a significant interaction between ambiguity and paragraph topic as questions about spatial paragraphs had long RTs regardless of ambiguity (see Figure 7). The interaction shows that spatial and ambiguous questions (respectively) take significantly longer to respond to than non-spatial and unambiguous questions. Further, it shows that there is a two-way interaction whereby questions about ambiguous, non-spatial situations take less time to process compared to unambiguous non-spatial ones than ambiguous spatial questions do when compared to unambiguous spatial questions. This indicates that spatial situations may increase the difficulty of detecting ambiguity.

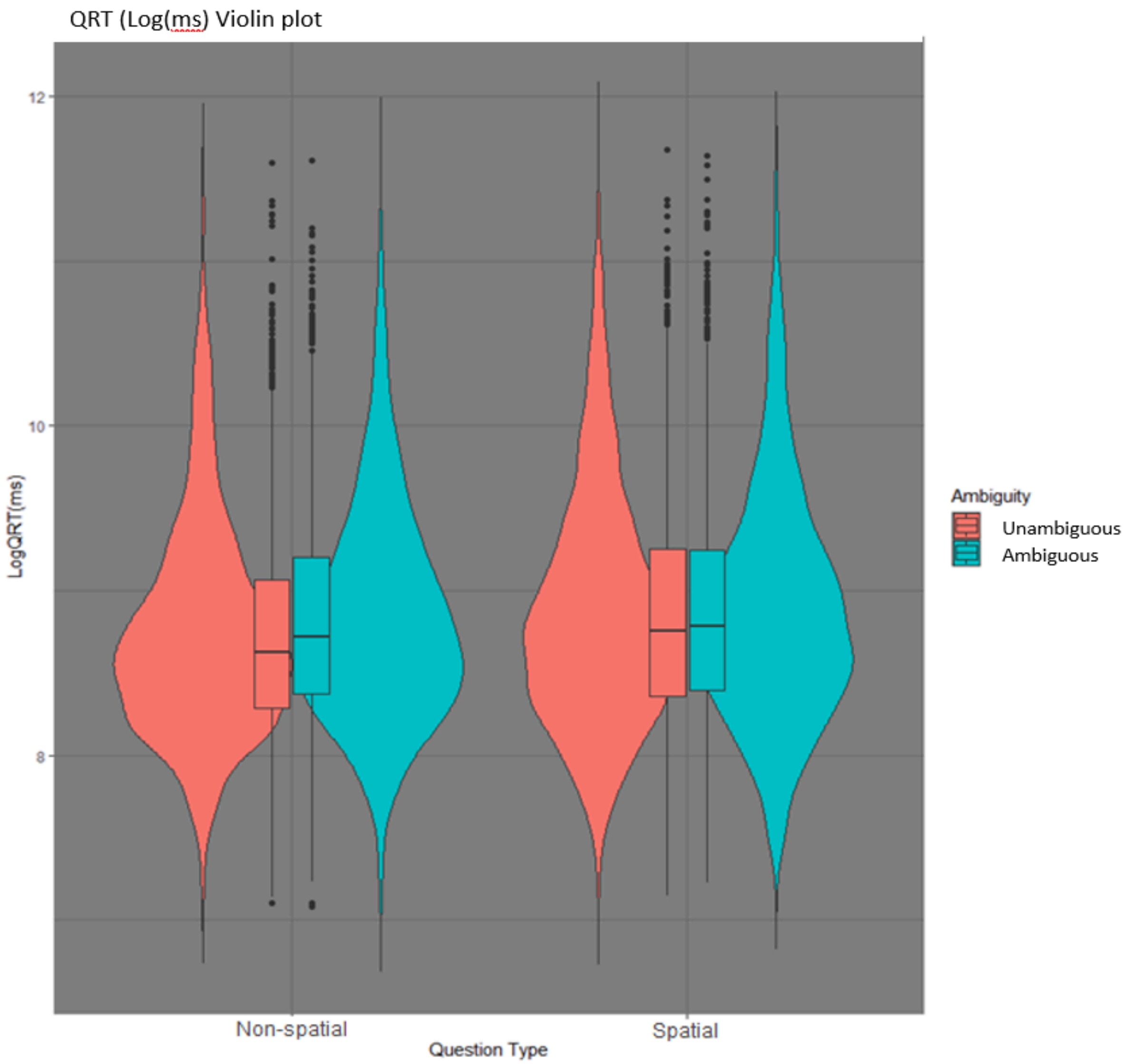
Table 3:

LMM Results for QRT log(ms) as a function of ambiguity and stimulus type (non-spatial/spatial).

Fixed effects	**log(QRT)**				
	b	**SE**	**df**	**t value**	**Pr(>|t|)**
Intercept	8.83	0.05	102.80	178.45	**< .001**
Spatial vs Non-spatial	0.04	0.01	6006.00	6.72	**<.001**
Ambiguous vs unambiguous	0.02	0.01	6007.00	3.05	**.002**
Ambiguous Spatial vs Ambiguous non-spatial	-0.02	0.01	6008.00	-3.43	**.001**

Figure 7.

Two-way interaction between Ambiguity and Question Type. The values were extracted using the “effects” R package v.4.1 (Fox & Hong, 2009) and Plotted using ggplot package for R (Wickham, 2016).



3.3 Question accuracy

We used GLMMs to explore participants’ accuracy when answering the situational comprehension question. As can be seen in Tables 4, 6, and 7, answers to questions about spatial situations were significantly less accurate than answers to questions about non-spatial ones. Further, questions about ambiguous situations were significantly less accurate than questions about unambiguous ones. However, we did not find any significant interactions between ambiguity and type. This may indicate that while spatial and ambiguous situations (respectively) are harder for participants to process than non-spatial and unambiguous situations, ambiguity detection in both spatial and non-spatial is “equally” difficult and may not compound the difficulty presented by processing a spatial situational model.

Table 4:

Mean accuracy by Type and ambiguity (SD in parenthesis)

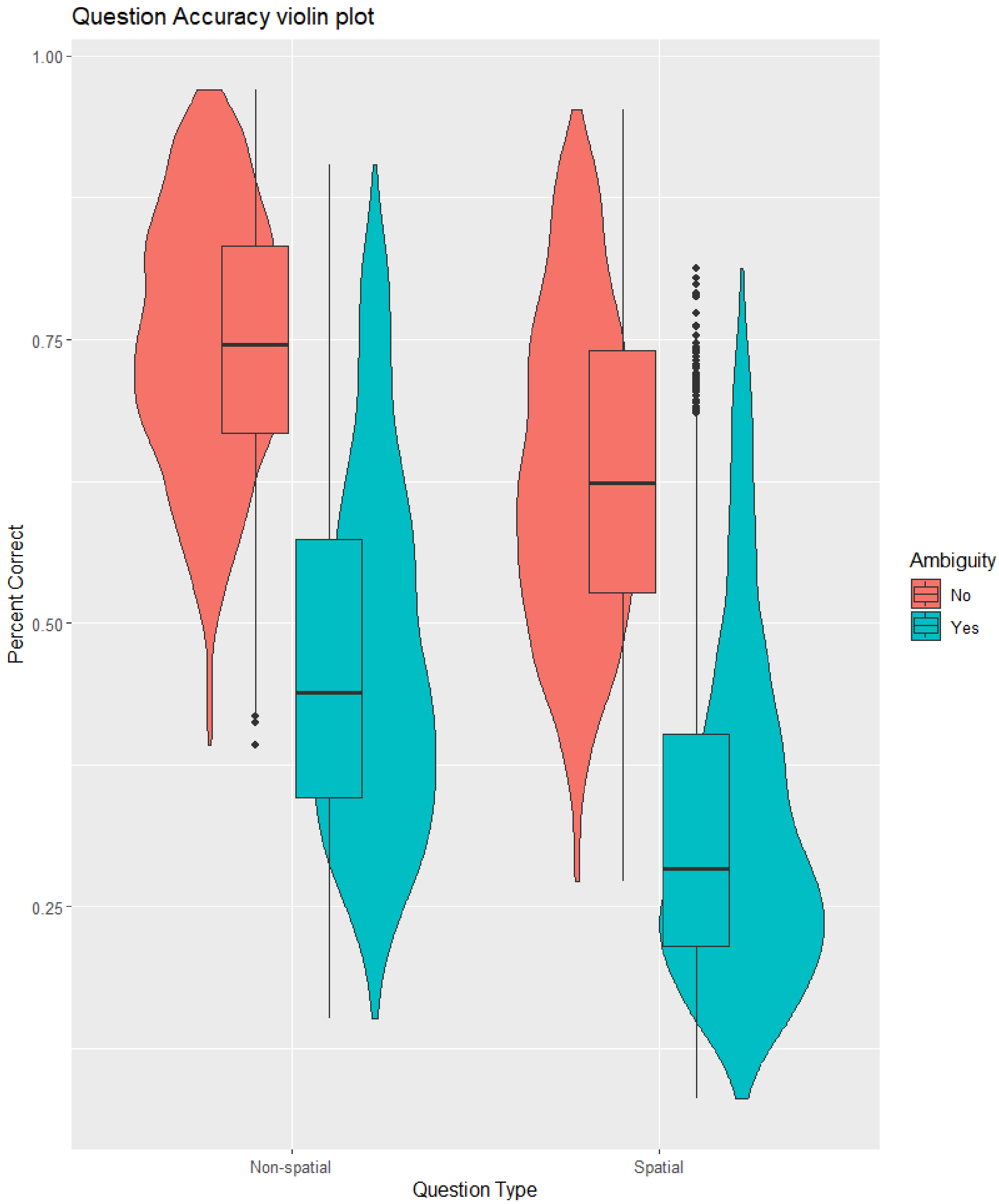
Ambiguity	**Type**	**Mean accuracy (%)**
Ambiguous	**non-Spatial**	0.541 (0.498)
Ambiguous	**Spatial**	0.334 (0.472)
Unambiguous	**non-Spatial**	0.791 (0.406)
Unambiguous	**Spatial**	0.669 (0.471)

Table 5:

GLMM Results for question accuracy as a function of ambiguity and stimulus type (non-spatial/spatial).

Fixed effects	**Question accuracy**				
	b	**SE**	**df**	**z value**	**Pr(>|t|)**
Intercept	0.22	0.08	-	2.61	**.009**
Spatial vs non-spatial	-0.31	0.03	-	-10.90	**<.001**
Ambiguous vs unambiguous	-0.69	0.03	-	-23.43	**<.001**
Spatial ambiguous vs non-spatial Ambiguous	-0.02	0.03	-	-0.85	.396

Figure 8:Two-way interaction between Ambiguity and Type influencing question accuracy. The values were extracted using the "effects" R package v.4.1 (Fox & Hong, 2009) and Plotted using ggplot package for R (Wickham, 2016).



3.4 Likelihood to select “There is not enough information”

We calculated the likelihood of selecting “there is not enough information” in the same way that “correct” answers were calculated, i.e. “If option 3 was selected “correct”, if not “incorrect” which applies to both ambiguous (where this was the correct option) and unambiguous questions (where it was not the correct option). We used a GLMM to explore when “There is not enough information” was selected. We did this to investigate if this option was used as an answer to ambiguous questions as per the instructions or as a “last resort” when participants could not find an answer. If used as a “last resort” option it would be likely that “there is not enough information” would be selected equally often for questions about ambiguous and unambiguous situations, along with an increase over QRT. As can be seen in Table 6 and Figure 9, there is a significant increase in the likelihood of selecting “there is not enough information” with increasing QRT. While this could indicate that this option was indeed used as a “last resort” (especially for unambiguous situations where this was incorrect) it could also indicate that ambiguity takes longer to detect. Figure 9 highlights a steep increase in the likelihood of selecting “there is not enough information” after the logQRT 10 mark, at least for the non-ambiguous, non-spatial situations. This may indicate a “last resort” selection, whereby a participant’s inability to answer the question causes them to eventually answer “there is not enough information”. With this said the likelihood of selecting “there is not enough information” is always higher for questions about ambiguous situations where the answer is appropriate indicating that participants can and do detect ambiguity, but this is at a chance level. We also found that the likelihood of selecting “there is not enough information” is significantly higher for questions about non-spatial situations than for spatial situations.

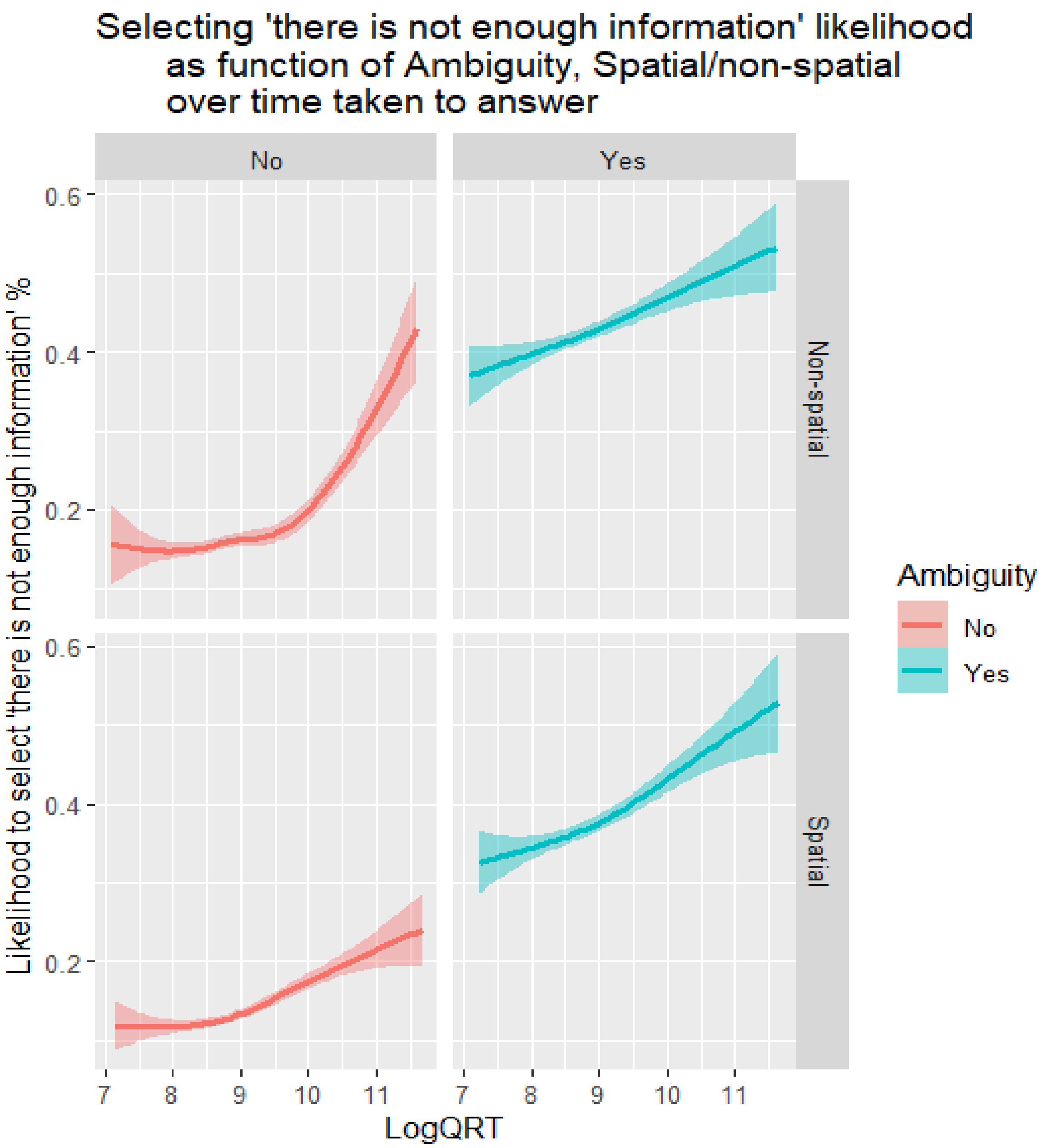
Table 6.

GLMM Results for likelihood to select "There is not enough information" as a function of ambiguity and stimulus type (non-spatial/spatial) and LogQRT.

Fixed effects	**Likelihood to select "There is not enough information"**				
	b	**SE**	**df**	**z value**	**Pr(>|t|)**
Intercept	-3.22	0.534	-	-6.03	**<.001**
LogQRT	0.22	0.059	-	3.80	**<.001**
Ambiguous vs unambiguous	1.47	0.067	-	21.98	**<.001**
Spatial vs non-spatial	-0.25	0.063	-	-4.01	**<.001**

Figure 9.

Predictive model showing the interaction between Spacial/non-spatial, Ambiguous/unambiguous, and QRT, influencing selecting “There is not enough information”, fitted using GAMM smooths



4.0 Discussion

We explored how spatial/ non-spatial ambiguity/non-ambiguity influences how a text is comprehended by readers. Specifically, we investigated if readers of text describing ambiguous situations could similarly detect ambiguity to locate an object/subject in unambiguous situations. This question about ambiguity detection was not previously addressed in mental model literature.

The key results are as follows: We found that ambiguous texts led to significantly increased QRT and decreased question accuracy. QRT was also longer for spatial than non-spatial situations. Further, the accuracies for both ambiguous and spatial situations were significantly lower than for non-ambiguous and non-spatial situations. Note, however, that there was no interaction in accuracy between ambiguity and type and the accuracy of questions, which implies that ambiguity is “equally” hard to detect in both spatial and non-spatial situations.

Given that questions about ambiguous situations take more time to respond to and that these have lower accuracy than unambiguous ones, it stands to reason that participants require more time to create a preferred mental model of the situation (Rauh et al., 2005) from which to make inferences. Participants make required inferences from one of the two viable models, which lead to a valid but incorrect answer as ambiguity was not detected. This may indicate that the preferred model is only created by the reader when the model is probed during questioning but this is hard to say for certain using the employed methodology, two-paragraph stimuli, and two subsequent questions. Alternatively, models which undergo a revision process based on newer information (model probing) often neglect non-preferred models (Rauh et al.,2005). However, the exact mechanism for how this “incorrect” model is arrived at or revised is unclear.

The significantly higher accuracy and lower QRT for unambiguous stimuli implies that unambiguous premises are simpler to reason with than ambiguous ones as these only have one model, not two. It may be that even though participants in this study were not asked to create a model "they prefer", they did so for both ambiguous and unambiguous situations, such a preferred model is far more likely to be "useful" in unambiguous situations.

Given that reasoners prefer to construct models "useful" to perform a task (Nejasmic et al., 2015), we speculate that; even if the stimulus is free of major semantic/syntactic "hints" that activate preferred models (Rauh et al., 2005) or "deactivate" others (Knauff & May, 2006), the reasoner still constructs a preferred model from which to answer questions. When creating a useful model, not only to detect ambiguity but also to make inferences about an unambiguous model, a reasoner may select one viable model of an ambiguous situation and make inferences based on this. Ergo not realizing that the situation has multiple viable solutions and ambiguity going unnoticed. A similar mechanism has been found in reading more generally. Though unrelated to mental models, readers often interpret an ambiguous text in a way that is only "good enough" (Ferreira et al., 2002) which allows them to arrive at a stable but necessarily valid interpretation from which to complete a task (Ferreira & Patson, 2007). When using "good enough" reasoning, ambiguity often also goes unresolved/undetected. We suggest that readers of ambiguous situations create a preferred model, which is stable (holds with the premises), from which to reason as this benefits the reader in making inferences about at least one model but does not avail them in ambiguity detection. However, the higher QRT for questions about ambiguous situations adds the additional caveat that the process by which the preferred model is created occurs during model probing. However, this is speculation as from this study it is not possible to determine the mechanism or at which stage a preferred model is created.

Reasoners end up preferring one interpretation over others when dealing with the object orders contained within models (Goodwin & Johnson-Laird, 2005; Nejasmic et al., 2015). An additional possibility as to why ambiguity often goes undetected may be that the reasoner holds on to the separate premises (e.g. "A right of B" and "C left of B") apart and does not combine them until a complete model is required for the task (Nejasmic et al., 2015). Using "gist" (Sachs, 1967) premises may be a useful strategy due to the memory load of having to remember two different situations, containing a total of four premises that describe six objects/subjects. Further, these may describe ambiguous configurations. With this in mind, creating a model incrementally (Johnson-Laird & Byrne, 1991) may not be beneficial for the reasoner when attempting to detect ambiguity as this could lead to only a single, preferred model being used to evaluate if the situation is ambiguous. Potential evidence for difficulty in integrating premises that create an ambiguous situation is shown by the significantly higher QRT for ambiguous situations, it may be that the process of integrating two premises that create an ambiguous order from memory takes significantly longer than integrating unambiguous premises. However, this is only speculation.

The situational ambiguities explored in this study are derived from two unambiguous premises which are put together in a manner that leaves relationships between entities underspecified. As such it may be more complicated for reasoners to combine these, especially when ambiguous. It is known that readers use different strategies to break down a text into smaller, more manageable chunks (Swets et al., 2007). Additionally, models are assembled from premises “bit by bit” (Johnson-Laird & Byrne, 1991), where the second premise is likely to be “inserted” into the first in a way that violates neither premise. This may take more time in ambiguous situations and allows for additional errors in reasoning which could result in a longer reasoning time for questions about ambiguous situations, lower accuracies, and a higher than normal likelihood to select “there is not enough information”. Further, it is known that readers have greater comprehension difficulty the greater the number of premises and entities in a model (Goodwin & Johnson-Laird, 2005). This may be due to object locations being added incrementally (Johnson-Laird & Byrne, 1991). Rather than storing the full model, known to happen in spatial reasoning (e.g., Lin & Matsumi, 2022) reasoners instead remember the premises(i.e “A right of B” and “ B left of C” rather than “A right of B, which is left of C”). The creation and later combination may influence how the reader creates and assesses the model as a whole. It is as of yet not known whether or not this influences a reader’s ability to detect ambiguity.

A longer question reading time may be the result of the difficulty in processing an ambiguous model, as these have more than one solution, but it is also possible that participants chose “There is not enough information” as a “last resort”. The likelihood of selecting “There is not enough information” increases over question reading time. However, this selection likelihood is significantly lower in the unambiguous condition compared to the ambiguous condition where it is the correct answer. The increase in selecting “there is not enough information” over time is especially prevalent in questions about unambiguous non-spatial situations, increasing sharply as QRT increases. As “there is not enough information” is not an appropriate response in unambiguous situations, it is likely that over time the use of “there is not enough information” is used as a “last resort”.

For questions about ambiguous situations, it is difficult to determine if the answer was used as intended or as a “last resort” answer. Participants may require longer to construct models from ambiguous situations and answer correctly rather than using a “last resort” or speed-accuracy trade-off. Furthermore, taking into account question reading time as a whole, these effects do not appear to be consistent with a speed-accuracy trade-off.

We also found that questions probing spatial relationships took significantly longer to answer than the non-spatial ones. Questions about spatial relations may be influenced by semantics (Knauff & Johnson-Laird 2002; Knauff & May, 2006) more so than non-spatial relations and are therefore more complicated for readers to interpret. Consequently, non-spatial situations may be “easier” to interpret than spatial ones as is shown by the higher model comprehension in both ambiguous and unambiguous non-spatial situations. Additionally, when recalling the situation during probing, irrelevant descriptives, e.g.,” It was a sunny day” may also interfere with the encoding of the premises “The car was left of the bike” (Dewar et al. 2007; Craig and Dewar, 2018), which may not happen with more abstract concepts such as wealth or popularity. This could be attributed to spatial models needing a different strategy, e.g., visuospatial strategies to visualise their order (Ford, 1995; Bacon et al., 2008). Non-spatial models may involve a verbal-propositional reasoning strategy that falls in line with the base interpretation of text which requires verbal-propositional reasoning to begin with (Handley et al., 2002). It is also known that reasoning strategies are often different from person to person (Bacon et al., 2003). Furthermore, it is known that spatial reasoning requires large amounts of working memory (Zwaan & Radvansky 1998; Friedman & Miyake, 2000; Gyselinck et al., 2007; Lin & Matsumi, 2022) and so having to remember two models, four premises (One model spatial the other non-spatial) may have caused additional strain on spatial reasoning ability.

A further difficulty in constructing the model may result from the reader visualizing the entire scene described in the paragraph rather than the situation alone, though whether this visualization of the described scene influences the non-spatial or spatial situational model is not yet known. It may be beneficial to use a single-paragraph paradigm in future work to reduce "semantic noise" from the other paragraph. In the case of this study, the reader needs to remember two situations, either one of which may be ambiguous, and process these in the order of the paragraphs without access to the text. A single paragraph could give clearer insight into how these models are processed as the reader is only required to recall one situation. Further interference in questions about spatial paragraphs may come from the reiteration of the perspective given in the stimulus. This perspective may have differed from the perspective from which a participant constructed the spatial situational model. It is known that reasoners use their own strategies for constructing and reasoning with mental models (Schaeken et al 2000). By outlining a particular perspective from which to reason (e.g left to right) a reasoner using a different strategy (e.g top to bottom) may have greater difficulty reasoning with the model, even if the model was originally a left to right description, the sequence/order of objects/subjects does not change. Alternatively, how the perspective is given in the stimulus is often through the use of some kind of "4th" object or subject, which may make spatial models more complex than non-spatial models (Goodwin & Johnson-Laird, 2005).

Questions probing non-spatial relationships do not require perspective to be answered correctly. For example, in a situation describing “Toby, Sarah, and Jacob” in terms of speed, the fastest will always be the fastest, but a situation where a “spoon” is left of a “fork” would change if viewed from the opposite side. Therefore, the increase in QRT for spatial questions may result from additional perspective-taking processes. For example, if the reader creates the spatial model using their preferred method (Schaeken et al 2000) or perspective e.g. Right-to-Left rather than the given Left-to-Right, then forcing the reader to answer from their non-preferred perspective may increase cognitive load. The higher QRT in questions about spatial vs. non-spatial situations may result from perspective being required to answer spatial but non-spatial questions. It is known that taking perspective plays an important role in processing object relations (Taylor & Tversky 1996) as well as descriptions of space (Levelt, 1996). This perspective may not be required when participants create a non-spatial situational model. For example, in a spatial situation the object furthest left becomes the object farthest to the right if viewed in mirrored form, but in a mirror “*Toby”* would still be faster than “*Jacob”*. Participants may use different strategies to compile and reason with spatial and non-spatial models, but it is not possible to determine this using the current methodology and data.

A potential, though more speculative, explanation for the differences in the processing of spatial and non-spatial situations is that the reader encounters difficulty when reading a described spatial situation due to semantic interference (Boudreau & Pigeau, 2001). As discussed in the introduction, spatial situations may be more influenced by semantics (See Example 4 & 5) (e.g., Castelain & van der Henst, 2021; Lopiccolo & Chang, 2021). Despite having been written to mitigate potential semantic interference in spatial situations, we still found significant differences in both accuracy and QRT between spatial and non-spatial situations. It is still not known if the strategies used to solve spatial and non-spatial situations are different but it is known that strategies differ between spatial and non-spatial syllogistic reasoning tasks (Ford, 1995; Bacon et al, 2008). Potentially, problems may arise if the same reasoning strategy is used for both spatial and non-spatial reasoning and may cause this difference in question accuracy but this is not yet known.

A further possible, but again speculative, explanation for the lower accuracy for questions about spatial situations is that readers not only use propositional logic to reason with described situations but also use mental images (Johnson-Laird, 2001; Tenbrink & Ragn, 2012). Mental images of spatial situations are "useful" as they can be rotated and moved and constructing a visual image of the spatial situation could aid a reader to make inferences from the model. For example, a mental image of "*Sarah* is in front of *Toby*" can be rotated so that "*Sarah* is behind *Toby*". However, this is less "useful" for non-spatial situations where propositional logic would be of greater benefit, for example, "*Toby"* is still faster than "Sarah" even if they have been "visually created" by the reader. The use of visual logic may be more cognitively costly than using propositional logic in a reasoning task, though this is unknown and beyond the scope of the study.

Lastly, in this study, we show that readers likely create preferred mental models and make inferences from them when these are embedded into larger bodies of text. While not surprising, this implies that readers specifically focus their attention on the premises. Furthermore, we show that readers are not particularly good at detecting ambiguity when embedded in larger bodies of text, but ambiguity detection has not been studied before, at least not in this way.

5.0 Conclusion

This work provides first insights into how people process and detect situational ambiguity. We established a novel experimental reading task to investigate situational model interpretation. Overall, this new task functioned as intended and we were able to replicate multiple prior findings about ambiguity processing.

In conclusion, this study indicates that ambiguity is much harder to detect than solving non-ambiguous situations. We argue that this is not because participants are incapable of detecting the ambiguity during stimulus reading but rather because a “good enough” representation of the text is a cognitively cheaper alternative but that results in errors reasoners are asked to detect ambiguity. Furthermore, we suggest that the difference we find in non-spatial and spatial comprehension likely results from different reasoning strategies. However, we find evidence for spatial relationships being harder to reason about than non-spatial ones. While this line of inquiry is novel, the findings in themselves are not. We admit that more research is needed to glean more information from this methodology. However, the method functions as intended and helps to build a strong foundation for future work investigating ambiguity detection in naturalistic reading.

Chapter 3: Implications of situational ambiguity and a general introduction to ageing

Reflecting on Chapter 2

Chapter 2 outlines numerous possibilities for why situational ambiguity may remain undetected. However, while the information gleaned from the study is useful, it does have several flaws. The main flaw is that of the method. By presenting two situational models simultaneously readers may begin to confuse the two premises of each model. Furthermore, it is known from research into mental models that reasoners experience difficulty based on how many entities are contained within a model (Goodwin and Johnson-Laird, 2005). It, therefore, stands to reason that readers of situational models would also experience difficulty when attempting to recall two models from text they no longer have access to. It is also highly likely that the increased working memory capacity load of two separate models influences text comprehension as a whole (Schurer et al., 2020).

Ambiguity, be it situational or otherwise, could also be influenced by a reader's history. Not just from a cultural perspective, where associations with words may influence how the reader interprets certain words, but also from what they have previously read. It is understood that readers maintain uncertainty arising from phonological or grammatical neighbors, which they had read previously, to sentences they read later (Levy et al., 2009). Here it was found that readers are aware of these ambiguities in text and use these to adjust their future expectations of what they read and this is directly influenced by the input provided previously. Speculatively, this may mean that a reader encountering an unambiguous situation in the first paragraph may consider the second paragraph to also be unambiguous. While this is uncertain if this is what occurred in Chapter 2 it may give a clearer picture to mitigate this through the use of only a single paragraph in future chapters.

Memory-based interference (Gordon, Hendrick & Johnson, 2001) arising from prior ambiguity may cause a later non-ambiguity to be more difficult to comprehend. However, more research is needed to know if this also applies to ambiguity arising from described situations as well. As the study presented in Chapter 2 did not make use of eye-tracking and two stimuli were presented simultaneously, it is difficult, if not impossible to assert whether or not ambiguity in a situational model may influence the parsing of a non-ambiguous model later on, (or the other way around).

The more elements contained within a model, the harder it is to reason out the model (Goodwin and Johnson-Laird, 2005). In our case, the two models were unrelated, though both had to be remembered for later questioning. This, further compounded by the possibility that a non-ambiguity may influence the parsing of a later ambiguity, as well as relying heavily on working memory may indicate that spatial models are not harder to comprehend or detect ambiguity in, in contrast to previous findings (Mani & Johnson-Laird, 1982). It may be the case that Spatial situational models and ambiguities are more influenced by an increased working memory load (from the two paragraphs containing 2 models from 4 premises). Furthermore, it is unknown if spatial and non-spatial models differ in their working memory load, though what is known is that certain words types, more fitting with their description may be recalled more easily than those that are not (Sonier et al., 2020), however, this is far beyond the scope of the study in Chapter 2 and beyond this thesis.

Concerning the possibility that spatial and non-spatial situational models are processed differently, it is very important to outline another major flaw of the pilot study presented in Chapter 2. As described in Chapter 2 there is a major issue relating to the spatial stimuli and questions. This issue comes not from the stimulus but rather from the question. The question probing the spatial part of the model usually reiterated the perspective e.g. “Mr.Roberts is standing at X facing Y, which way must he go to get to Z” as perspective was deemed to be important for participants to be able to solve the model/detect ambiguity accurately. However, this introduced a 4th object/subject into the model (From now on referred to as a “4th Actor”). As is known from (Goodwin and Johnson-Laird, 2005) the more premises or objects are included in a model the harder it is for participants to solve the model. It stands to reason that the introduction of a 4th “actor” would also decrease the likelihood for participants to detect ambiguity. As it stands, the assertion made in chapter 2 that described situations relating to space are harder to comprehend than non-spatial ones may be untrue or at the least incomplete. However, we remedy this in the next chapter and further address this issue in Chapter 6.

Lastly, we found significant differences in the question accuracy of spatial and non-spatial subject types whereby spatial situational models appeared to be harder for participants to answer than non-spatial ones. In the study, we argue that a potential reason for the difference between spatial and non-spatial comprehension is due to the axis on which the model is created and how a non-spatial model may be oriented in any way that is preferable to the reader. By comparison, the spatial model is harder because it must be built on the same axis in which it is outlined i.e. left to right, top to bottom, etc. While it is of course possible to create a spatial model in any axis one wishes, solving the model would require the model to be rotated back into the original axis. Naturally, this process adds steps that may interfere with reasoning. Therefore we postulate that the reason for the significantly lower spatial comprehension in spatial models is due to one of two reasons:

1. The model can’t be built in a reader's preferred reasoning axis and is therefore harder for them to comprehend.

Or

2. The model is built in the reader's preferred axis but must then be rotated to be reasoned with during questioning.

At this point, it is not possible to establish a clear answer as to what may be happening. The study in Chapter 2 does not separate the two paragraphs and as such it is not possible to interpret the stimulus reading time in a meaningful way for this purpose. However, if we assume that the model can't be built in a reader's preferred reasoning axis we should see a significantly higher SRT for the spatial stimulus, as this stimulus cannot be encoded in a reasoner's preferred axis to begin with. Further, we might see a significantly higher QRT as readers struggle with the reasoning in a non-preferred axis. On the other hand, if the model is encoded in a reader's preferred axis during stimulus reading we may not see a difference in SRT but a difference only in QRT, as this requires the rotation of the model for reasoning to happen accurately. While we currently lack any significant evidence to support such a hypothesis it is possible to find clues from research in the field of Synesthesia. Firstly, it is known that sequence/space synesthesia is relatively common (Ward et al., 2018) and is estimated to influence 8- 12% of people. If we take Ward et al's findings at face value then it is highly likely that a solid chunk (8-12%) of our participants were influenced by this. While the study in question used number orders, rather than described orders of objects, etc, it is still possible that their findings are relevant to us and this thesis. Furthermore, it has been found that even persons without synesthesia have preferred, a variety of methods when ordering objects (Rizza & Price, 2012; Havlik et al., 2015). However, whether or not this is relevant in this research is as of yet not known.

In more specific terms discourse processing may offer some explanations as to how texts describing social and spatial situations are interpreted. Firstly, in general, reading working memory consists of several dissociable components (Roberts and Gibson, 2002), in the case of mental models these components could be made up of the individual premises. Additionally, Roberts and Gibson (Roberts & Gibson, 2002) describe the clause as an important part of a sentence. This would serve to help to explain how a single clause premise in a mental model like a reading task is remembered and processed in that each premise represents a single clause, which it usually does, and as such is remembered as a single relational “chunk” which can then be placed aside any other premise to create a model of the situation described in the stimuli text. It is also possible that syntactic order plays a key role in the processing of a premise whereby a participant’s comprehension of object-relatives is based on the syntax of the premise. For example, capacity (Andrews et al., 2006) has shown that comprehension of object-relatives relied more heavily on a domain-general capacity to process complex relations than on working memory capacity. The study used short simple sentences and presented them with different syntaxes that did not change the meaning of the sentence. This is something that can also be done in spatial and social order premises with the added caveat that such a syntactic change might result in the premise being interpreted as ambiguous.

If one were to draw the comparison of a mental model premise, e.g “ A right of B” to that of an object pairing it is apt to also consider the findings of (Larkin & Burns, 1977) where participants were given a list of stimuli and were asked to recall them in pairs. Participants paired lists of four different lengths 4, 6, 8, and 10, here there was an overall effect of embedding whereby embedding was significant in all conditions. Larkin and Burns indicate that this argues against linguistic difficulty being a hindrance to participants. If we assume that these word pairs are similar to social or spatial premises it would then stand to reason that so long as the direct relationship between the two objects in that premise is not syntactically unclear, nor obscured, that re-reading of this premise plays a significant role in correctly encoding the object pair. With that said it is also the case that the relative syntax of each relative clause containing a pairing plays a large role in comprehension (Scontras, Badecker, and Fedorenko, 2017). Though this is less relevant to mental model-like premises, it should also be taken into consideration should the syntax of an ambiguous premise result in an overly complex change to the order of words. How the relationship in these object pairs is interpreted is as of yet somewhat unclear though it appears that the temporal order of relationships in text could result in participants interpreting such a premise.

A good way of figuring out the role played by the premises would involve the use of eye-tracking while participants read such premises. Eye-Tracking has been used in multiple reading experiments, though has not often been used to investigate preferred mental models. In the next chapter, we aim to shed some light on how reading patterns differ between spatial/non-spatial, ambiguous, and unambiguous premises. We also further investigate the effects of ageing on eye movements when reading situational models.

In conclusion, the evidence presented in Chapter 2, while useful, raises more questions than it answers. This is especially true when considering the aforementioned working memory load and populations in which working memory is reduced. Something that awaits us all as we age.

Ageing

Numerous cognitive deficiencies begin to manifest themselves as we age, this is true for our ability to navigate as well as our ability to read and comprehend texts. In the case of the latter, we begin to develop strategies to compensate for or mitigate these deficits. For example, we may begin to take more risks when reading (Rayner et al., 2006), whereby older readers skip more words than younger readers might. This risky reading strategy is thought to be the result of saving time while reading due to age-related slowing of reading. However, what has not yet been investigated is how ageing affects the interpretation and understanding of situations described in the text. Nor is it known to what extent the risky reading hypothesis applies to the comprehension of the text. Especially described situations from which the reader must reason or determine ambiguity.

The risky reading hypothesis suggests that relative to younger readers, older readers tend to make more and longer fixations as well as longer saccades (Paterson, McGowan & Jordan, 2013), and they are also more likely to skip words and regress to prior words more frequently. It is thought that older readers “adopt” this risky reading strategy to compensate for slower rates of lexical processing and use context to “guess” words. However, this could also be a strategy to compensate for a smaller less symmetrical perceptual span (McGowan & Reichle, 2018). As such, an older reader, more inclined to skip words to begin with, may miss important information in a text, especially if we consider that more predictable words are skipped more often than non-predictable ones (Rayner 1998, Choi & Gordon, 2013). This may mean that older readers are more likely to incorrectly assume specific content in an ambiguous text as such misinterpret it. However, it is not yet known if this is also what happens in situational ambiguity. It should be noted that Chinese readers (reading Chinese) seem to have adopted the reverse strategy--performing shorter saccades and skipping words less often (Wang et al., 2016; Zang et al., 2016).

Importantly, the precise reasons for older readers' risky reading are not fully understood, nor is it known at what age this strategy becomes more prevalent. While the risky reading strategy is an important aspect of the investigation of reading in ageing, it is also pertinent to take into account that oculomotor deficits may be at play in differences between older and younger readers. For example, while certain aspects of oculomotor control such as the visual-vestibular-ocular-reflex decline in age, others, more important to reading, do not (Kerber, Ishiyama & Baloh, 2006). It is also known that in saccade tasks age does not affect accuracy or binocular coordination (Yang & Kapoula, 2008). However, physiological characteristics do show differences as it has been found that older readers have a smaller, more symmetrical, perceptual span (Rayner, Castelhano & Yang, 2009). This has been explored in relation to the aforementioned risky reading strategy and seemingly links the different perceptual span in older readers to the adoption of the risky reading strategy, allowing older readers to maintain a similar reading speed to younger readers (Risse & Kliegl, 2011). Therefore allowing a reader to read at their own pace, re-reading as much as they want, could negate deficits attributed to risky reading, whether they come from oculomotor deficits or not, instead allowing for a more in-depth investigation into cognitive deficits. The fact that this thesis uses a naturalistic, self-paced reading approach may prove useful to further understanding the risky reading behaviours of older readers, in terms of how eye movement patterns of natural reading influence the understanding, or as it may be a misunderstanding of situational models.

However, the process of reading itself changes with age, as highlighted by Paterson et al. (2013) in their investigation of how spatial frequency sensitivity affects reading performance in young and older adults. While both age groups exhibit good overall reading ability and comprehension, they rely on different spatial information to support reading. Age-related changes in eye movements also influence reading behaviour, as shown by Won Jun Lee et al., (2019) and Dowiasch et al., (2015). Lee et al., (2019) found that the range of eye movement decreases with age, particularly in long horizontal and upward gazes, which may impact reading efficiency. For example, an older person attempting to re-read a part of a previously read text (upward of current eye-position) may struggle to accurately launch and land the saccade in the desired location. Further, Dowiasch et al.'s research in real-world settings suggests that while ageing affects saccade parameters, it may not significantly impair tracking eye movements during tasks such as walking, suggesting that perhaps saccades during reading are “special”, though this assertion is beyond the scope of this Thesis. However, these findings underscore the complex interplay between ageing, eye movements, and reading, highlighting the need for tailored interventions and accommodations to support reading and cognitive health in older adults. Research conducted by Chang et al (2021) suggests that reading frequently can help mitigate cognitive decline in older adults.

Furthermore, it’s also known that older readers tend to hold on to incorrect assumptions about ambiguous texts such as garden path sentences (Christianson, Williams, Zacks & Ferreira, 2006), even when later prompted to the correct interpretation of the text (Ferreira et al., 2002). While the good enough interpretation of garden path sentences is prevalent in populations of all ages it also applies to non-garden path sentences as it is highly task-dependent (Ferreira and Patson, 2007). Therefore, it is possible that the same effects shown for good enough interpretation in older readers also occur in situational model reading and may be exacerbated due to ageing. While such an effect in situational models has not yet been found it may stand to reason that older readers may maintain a preferred, stable, version of the model they have just read, rather than two possibilities which may entail using a larger amount of memory.

An additional ageing influence on situational model interpretation could come from the perspective used to describe the model, which does not differ from stimulus to question (See Chapter 2 Methods). However, as is known from work on mental models, reasoners use a variety of strategies to reason out a model (Schaeken et al 2000). Therefore it is possible that the perspective iterated in the stimulus and reiterated in the question does not conform with the reasoner's personal preferred reasoning strategy and therefore causes difficulty when it is not needed. This is something that becomes especially poignant concerning ageing. Numerous studies have found that older populations are more likely to struggle with navigational and perspective tasks. It is known for example that the ability to detect changes in object locations decreases with age (Muffato et al., 2019), as such it may also be the case that such deficits translate into object locations in the written word. Rather than swapping objects in their locations this study instead investigates objects that have no fixed position (ambiguous) whereby older readers may have decreased sensitivity to detecting this ambiguity. Furthermore, social situational models, which hold objects not in space but rather in a ranked order, such as ability or popularity may also be affected in the same way, though this is also not yet certain.

It is with these uncertainties raised by Chapter 2, and subsequent discussion that we must now move on to Chapter 4. This chapter attempts to correct the mistakes of Chapter 2’s study as well as delve deeper into the age and situational models. However, chapter 4 does not yet address the problem of the introduction of the “4th actor” in spatial questions. Rather chapter 4 serves as a baseline that investigates the reading of described situations for us to determine if older people read described situations differently and if so why/when they might do this.

Chapter 4: Ageing and eye-movements reading Situational Ambiguities

In keeping with open science, the full reproducible analysis code, raw data and supplementary materials are available on the OSF link provided below.

Abstract:

In Chapter 2 we found that people are not good at detecting ambiguity in underspecified situations. In this chapter we used eye-tracking in order to better understand how premise information is encoded and to gain further insight into when a mental model is created by the reader. Furthermore, we also began to investigate the potential differences between younger and older readers paying close attention to known behaviours of older readers. For example the use of a "risky reading" strategy (Rayner et al., 2006). The stimuli and question type were the same as in Chapter 2 but used a single paragraph. In line with predictions from the risky reading strategy, we found that older readers skipped words significantly more often and re-read more often than younger readers. However, we also found that younger readers are significantly more likely to re-read ambiguous stimuli which may indicate the adoption of a task dependent strategy and that model creation may occur during stimulus reading rather than from memory during probing. We found that both older and younger readers were significantly more likely to re-read "premise words" (words describing relationships between objects/subjects). However,we admit there are problems with this assertion as older people re-reading cannot be "separated" from re-reading as a result of the risky reading strategy. We speculate on how working memory compensation strategies such as "risky reading" may play a role in ambiguity detection and suggest that both older and younger readers may adopt different strategies in order to better make inferences about their mental models.

Introduction:

Improving methodology:

In Chapter 2 we found that readers of situational models, consisting of two premises, which were embedded in text could accurately make inferences about both spatial and non-spatial situations. However, readers were significantly less able to detect when such a situation was ambiguous. We speculated that the two premises were only being integrated during probing which led to a higher Question reading/reaction time for ambiguous situations. However, as the experiment in Chapter 2 made use of two stimulus paragraphs, one spatial one non-spatial, it was difficult to determine if this was the case. Furthermore, this methodology also made it hard to find clear evidence for alternative mechanisms of model integration and finding the cause for low ambiguity detection.

We also found that spatial situations were seemingly harder to comprehend than non-spatial situations for both unambiguous model comprehension and ambiguity detection. We speculated that the reason for this was that spatial situations may be more influenced by semantic cues (Knauff & Johnson-Laird 2002;Knauff & May, 2006) or problems with scene visualisation (Ford, 1995; Bacon et al., 2008). However,again the two paragraph stimulus used in the prior study made this difficult to accurately speculate upon. We couldn’t be sure as to what interference was coming from where. We suggested also that a participant having to remember two models, four premises and the order of 6 objects/subjects may also have caused an additional working memory load. Making it difficult to determine, or accurately speculate on, if different mechanisms/problems are involved in the processing of spatial and non-spatial models.

The lack of an accurate measure of stimulus reading time, resulting from a two paragraph paradigm, makes it harder to determine how readers create models from which to reason and detect ambiguity. Further, the use of two paragraphs makes it harder for readers to perform the task required of them correctly. Therefore, in this Chapter's study we remedy this by presenting only one paragraph of stimulus text with one subsequent model probing/ambiguity detection task. Furthermore, the study made use of eye-tracking in order to more precisely measure the reading of the stimulus paragraph and allowed us to investigate how the premises are read.

As we saw in Chapter 2, a situation which describes the relationship between three objects/subjects can create an ambiguous situation where there are more than two viable orders of objects/subjects. In typical mental model research the reader is not asked to "detect" ambiguity, i.e. state if a situation has more than one valid interpretation. Research into mental models investigates the creation of a "model" in one's mind from the relationship between objects/subjects (Franz & Garnham, 1990; Baddeley, 1998; Johnson-Laird & Byrne, 1991; Rauh et al., 2005; Goodwin & Johnson-Laird, 2005; Knauff & May, 2006).

In Chapter 2 we established that the detection of ambiguity is harder for reasoners than selecting the correct answer in an unambiguous situation. However, while the detection of ambiguity was significantly harder than answering questions about an unambiguous model for reasoners, they were certainly capable of doing so some of the time and we suggested a few reasons for this. However, because the two paragraphs were presented simultaneously, we couldn't investigate the stimulus reading times to understand why the ambiguity was not always detected . Chapter 2's two paragraph presentation made it difficult to investigate the overall reading time of each stimulus. As a result it was not possible to discern if ambiguous, unambiguous, spatial and non spatial paragraphs took more or less time to read. The use of eye-tracking allows us to remedy this shortfall and expand upon it by making it possible to investigate how premises are read in comparison to non-premise text.

Furthermore, Chapter 2 found that spatial situations are harder to interpret than non-spatial ones. We suggested that this may be due to higher semantic biases being present in spatial situations as these are more likely to have to adhere to “real world” rules, whereas non-spatial situations may not. However,due to an additional complication in the way in which spatial questions were probed, using a “4th object/ actor”, we also cannot be sure if spatial situations really are harder to interpret. We also remedy this problem in this study, by asking about the “middle” of the model. The “middle” of the model refers to the object/ subject which is “between” the two others, for example if premises have described a situation where “A is right of C and B is left of A”, we ask participants to determine “which of the three objects is between the two others”. This style of questioning removes any need for additional perspective and allows the questions between spatial and no-spatial to be more comparable as they can now be worded almost identically. Therefore in this study it is likely that we will find the prior significant difference in comprehension accuracy between spatial and nonspatial are no longer present.

Mental models are not typically studied through the use of eye-tracking, although this has previously been investigated for different purposes (Sima et al., 2013). Therefore this Chapter addresses the question as to whether or not eye-movements change based on whether or not the described situation was ambiguous or not. It is currently unknown as to how eye-movements and reading behaviour, already changed by age, are influenced by the presence of ambiguity in the reading of situational models; it may be the case that eye-movements influence the construction/ rejection of situational models.

While eye-tracking may be useful to better understand how premises are read and understood, it is important to consider that this study also tests how older readers detect ambiguity and if this is different from younger readers. This study functions as a “pilot” for testing older readers in terms of their ability to detect ambiguity and reason with mental models embedded in text. It is therefore important to understand some of the fundamental differences between how older and younger people read.

Reading, eye-movements and ageing:

It has been shown that eye movements during reading differ between age groups; for example, children tend to make more and longer fixations and generally make shorter saccades (Blythe & Joseph 2011; Parker, Slattery, & Kirkby, 2019). Considering that there is a difference in reading behaviours in children compared to adults, it would stand to reason that there are behavioural changes that also occur in normal ageing. This is something that has been explored before, notably resulting in the formulation of the risky reading hypothesis (Rayner, Stroud, Reichle, Pollatsek, & Williams, 2006). The risky reading hypothesis suggests that, relative to younger readers, older readers tend to make more and longer fixations as well as longer saccades (Paterson, McGowan & Jordan, 2013), and they are also more likely to skip words and regress to prior words more frequently. It is thought that older readers "adopt" this risky reading strategy in order to compensate for slower rates of lexical processing and use context in order to "guess" words. Though this could also be a strategy to compensate for a smaller less symmetrical perceptual span (McGowan & Reichle, 2018). Though it should be noted that this does not seem to hold true for observations involving older readers of Chinese who seem to have adopted the reverse strategy--performing shorter saccades and skipping words less often (Wang et al., 2016; Zang et al., 2016).

However, the precise reasons for older reader's risky reading is not fully understood, nor is it known at what age this strategy becomes more prevalent. While the risky reading strategy is an important aspect to the investigation of reading in ageing, it is also pertinent to take into account that oculomotor deficits may be at play in differences between older and younger readers. For example, while certain aspects of oculomotor control such as the visual-vestibular-ocular-reflex decline in age, others, more important to reading, do not (Kerber et al., 2006). It is also known that in saccade tasks age does not affect accuracy nor binocular coordination (Yang & Kapoula, 2008). However, physiological characteristics do show differences as it has been found that older readers have a smaller, more symmetrical, perceptual span (Rayner et al., 2009). This has been explored in relation to the aforementioned risky reading strategy and seemingly links the different perceptual span in older readers to the adoption of the risky reading strategy, allowing older readers to maintain a similar reading speed to younger readers (Risse & Kliegl, 2011).

While at first glance it appears that the body of research on saccades in ageing differs widely in conclusions as to whether saccadic accuracy changes in older adults or not, this is more complex than it at first appears to be. For example (Warabi et al., 1984) found that older adults' horizontal saccades along 40 degrees did not differ significantly from those of younger adults, although older adults (n=24) showed a significantly greater latency compared to younger adults (n=8), something found in almost all other senescent saccade studies. While the older group differed in their mean saccadic latency by 100ms from the younger group. In terms of saccades, the normal older group's saccadic accuracy, amplitude and velocity did not differ from that of younger participants until a saccade above 20 degrees of visual angle was performed. Approaching saccades of 40 degrees, the accuracy, amplitude and velocity of the older group began to plateau compared to the younger group. In both the normal older and younger groups, corrective saccades did not differ greatly in their latency (between 130-230 ms).

On the other hand, papers using a similar experimental paradigm report different results. For example Sharpe and Zackon (1987) found significant differences in the saccadic accuracy of their older participants (66-87 years) compared to their middle-aged group (35-63 years), which undershot more and more as the distances of targets increased.

However, no significant differences were found in saccades to 5-degree targets, which may not be too dissimilar to a large intra-line saccade in reading. Additionally, it was also found that there was a comparatively greater deviation in larger saccades, beyond 20 degrees. Similarly to Warabi, Kase & Kato, it was also found that older readers had significantly longer latencies than the younger and the middle-age groups. It should be noted that in both studies participants had to make saccades of unpredictable amplitudes due to the target appearing at an unpredictable location along the horizontal axis. While the differences found within saccade latencies should hold true in a reading task, the ones patterning to saccadic accuracy may not fully reflect on return sweep saccades nor saccades in reading in general as within text saccades tend to be guided by the predictable presence of text.

Overall, it appears that older people have a tendency to struggle to varying degrees when tasked with performing long saccades, at least relative to their younger counterparts. In studies that looked at shorter range saccades it appears that older people don’t differ significantly from younger people in terms of saccadic accuracy. For example, Yang and Kapoula (2008) showed this using a vertical saccade paradigm where the maximum saccade size participants were tasked with was only 15 degrees. In keeping with previous work, it was reported that older participants had significantly longer latencies than the younger participants did. Therefore, it appears that differences in saccadic accuracy are down to age but only when taking the length of the saccade into account, the same could also be said for the increase in hypo metric saccades as shown in (Irving et al., 2006). Though it does appear that saccadic latencies increase as we age, it is not yet certain if these differences are further exacerbated by saccade length.

While eye-movement control during reading has been investigated in older populations before, no studies have directly examined how eye movements in older readers are influenced by reading premises from which to create mental models. Additionally little is known as to how normal ageing influences the return sweep in reading. Although the return sweep is not the main investigative goal of this study, nor a secondary one, it is important to note that prior research on the return-sweep may shed light on eye-movements when re-reading and how this is influenced by ageing and the reading of mental model premises. Return-sweep saccades move readers' gaze from the end of one line to the beginning of the next one (Rayner, 2009). It has been shown that the return sweep is influenced by a number of text based factors such as font size and line length (Vasilev et al., 2019) whereby a larger font resulted in return sweeps landing further to the right of the margin compared to a smaller font. Additionally other factors such bolding of text affects return sweep accuracy, for example (Slattery & Vasilev, 2019) where bolding the line initial word resulted in return sweep saccades that required fewer corrective saccades (leftward saccades immediately following a return sweep) than in a non bolded condition indicating that bolding possibly reduces oculomotor error in return sweeps. Though it is still unknown as to how these factors impact the return sweep in ageing. It may be possible that some changes occur as it has been shown that older people may be more susceptible to distractions in saccade tasks (e.g., Bowling et al., 2014) which could show up in the return sweeps of older readers. Furthermore, distractions and inaccuracies in saccade planning may interfere in an older reader's ability to "search for" premises accurately, in order to re-read premises to better construct a mental model.

Additionally, it is not known how exactly saccadic range error in reading is affected by age. Should it be the case that older readers perform similarly to younger readers in return sweeps, it would be logical to assume that return sweep saccades do not change with age. However,as outlined by Warabi et al., (1984), and previously discussed, older readers have a significantly higher latency when "planning" and "performing" their saccades. This latency of older readers must therefore be taken into account when interpreting the reading of premises, so as not to confuse behaviour occurring as a natural result of ageing with behaviour as a result of reading premises (and constructing mental models). Age related saccadic range error and saccadic latency is also known from more recent non-reading tasks e.g (Yang & Kapoula, 2008). However, return-sweep saccades may be influenced by reading strategies such as the so-called "risky reading strategy" (Rayner et al., 2006). If so, then older readers may be more likely to target return-sweeps further into the next line (i.e., further from the left margin) and then subsequently require a regression to an earlier location on the line.

It stands to reason that difficulties in text interpretation (e.g., forming complex mental models) may also influence saccadic latencies and saccadic accuracy. Though there is ample research into how senescence influences certain ambiguity resolutions (e.g in garden path sentences: (Christianson et al., 2006; Yoo & Dickey, 2017), it is not yet known to what these may influence eye movements. However,it is well established that working memory declines as we age (Bopp & Verhaeghen, 2007; Borella et al., 2008; De Beni et al., 2004; Fiore et al., 2012 ; Park et al., 2002). It is known that visuo-spatial working memory is inherently linked to spatial reasoning tasks (Wang et al., 2018) and so an age related decline may negatively influence a person's ability to reason in space. Not only would this influence the interpretation of the text but might also influence the planning of saccades, However,this is conjecture.

However, age related changes in text interpretation are known. For example, research indicates that readers with a lower working memory span use different strategies to break down a text into smaller, more manageable chunks (Swets et al., 2007). Therefore, an age-related decline in working memory may influence the processing of situational ambiguity. For instance, memory has been shown to influence the parsing of long-distance dependencies (Nicenboim et al. 2015), where it was found that low capacity readers make more regressions, something that is already a known behaviour of older readers (Li et al., 2019; Rayner et., 2006). Additionally, processing speed slows with age (Myerson et al., 1990), which may influence accuracy or preference in time dependent tasks. Therefore, older readers should be less aware of situational ambiguities, which require remembering multiple premises, thus forming a large amount of information to be processed in one go.

Little is known as to how age influences either the processing or encoding of situational models. However, it is known that reading comprehension (Hannon & Daneman, 2009) does decline in age but that this is related to a general decline in working memory. Conversely, it has been found that there are no significant age related differences in making correct inferences from text (McKoon & Ratclif, 2013) and that younger readers are more likely than older readers to "hold on" to their prior inferences, even when these are later proven to be untrue (Guillory & Geraci, 2010). There also appear to be no differences in competence in making inferences about mental models (e.g.,Light, 1988; Radvansky et al.,1990). However, it is known that older people may struggle to integrate premises into a model (Copeland & Radvansky, 2007). Though, if these findings hold true when detecting ambiguity in mental models is not yet known.

Previous findings highlight age-related declines in reading speed (Kemper et al., 1993; Liu et al., 2017), but a slow reader is by no means an inaccurate one. However, older readers are more prone to skipping words (Rayner et al., 2006), thought to be implemented by the older reader as a "time saving" mechanism. Somewhat paradoxically, they are more likely to regress words that have previously been skipped causing an increase in overall reading time.Older readers may be "hedging their bets" on how much of the text they can skip based on their (presumed) greater knowledge of the world and reading experience. It is thought that older readers "adopt" a risky reading strategy in order to compensate for slower rates of lexical processing and use context in order to "guess" words (McGowan & Reichle, 2018). However,it is not known how the risky strategy might influence the reading of individual premises that are required in order to create the situational model as a whole.

A risky reading strategy used in the reading of texts which contain important, single word information may influence the model which is created from the combination of the two premises. For example, if an important spatial premise word such as "left" is skipped by the reader, they may struggle to accurately create the model. Alternatively, if a word is skipped and later regressed to, and/or re-read multiple times then despite an initial risky read, encoding the premise may be improved. Furthermore, it is possible that a regression or re-reading heavy based reading strategy is also adopted by younger readers as they learn the importance of premise encoding throughout the experiment. However, it may be difficult to differentiate the reading behaviour “adopted” by an older reader, attempting to encode premise information, from that of “risky reading”.

Re-reading is likely a great boon to the text comprehension of older and younger readers, as frequent encoding cements the important parts of the text (premises) in the readers working memory, making later retrieval easier(Margolin & Snyder, 2017; Rawson, et.al ,.2000; Stine-Morrow et.al ,. 2004). However, due to “risky reading” (Rayner et.al., 2006) and lower working memory capacity (Bopp & Verhaeghen, 2007; Borella et al., 2008; De Beni et al., 2004; Fiore et al., 2012 ; Park et al., 2002), We predict that eye-tracking will likely show that regressions are significantly more prevalent in older readers than in younger readers. Furthermore, in a text such as:

“Sally knew that she was smarter than Robert and she was sure that she would take first prize”

We predict that we would see higher levels of regression, and perhaps longer fixation durations for specific words in the text such as:

*“**Sally** knew that she was **smarter** than **Robert** and she was sure that she would take first prize”*

The text outlines a situation in which “Sally is smarter than Robert” in the initial half of the sentence. If the reader is aware that they will be asked about the situation described in the text then it is likely that the words describing the situation will be re-read more often than the ones that do not. From this point onwards we will refer to these words, such as; names, objects, left of, right of, smarter, poorer, in the premise as premise words (PW). Additionally it is important to note that what we define as “premise words” are merely “directional” words contained within the premise (e.g "left, right, better, worse, smarter, etc"), which determine the relationship from one object/subject to another. For a full list of premise words used please see the OSF for this thesis (found in the Stimuli folder and labelled "Chapter 4 premise words").

It has been suggested that when a text remains available to the reader that the position of specific target words is stored rather than the meaning of the target word itself (Inhoff, et.al., 2019). This could be useful to the reader when building a model of a situation as an object/subject within a premise can be "checked ", i.e. regressed to and re-read, prior to constructing its relation to another object/subject within a premise. In our case the reader would return frequently to these "target" premise words. If an older reader, more prone to skipping through risky reading, "misses" an important part of a premise, it may be the case that the general location of the premise in the text is stored and can be accessed more easily upon re-reading. While this is speculation, it would also support a prediction for higher rates of re-reading in older readers, especially for the premises. It is known that re-reading is a useful tool to improve reading comprehension (Wong & Moss, 2021), and this is not limited to single word re-reading; re-reading a text significantly improves reading comprehension (Margolin & Snyder, 2017; Rawson, et.al., 2000).

As working memory capacity influences text comprehension (Schurer et al., 2020) then it is useful for both younger and older readers. A reader with a lower working memory capacity, may, for instance, need to "filter" out unimportant parts of the text and create a "gist" level understanding (Sachs, 1967) of the text. In the case of reading a text from this study, the reader could/should/might focus their attention on the premises which describe the spatial or non-spatial situation. Re-reading the premises may act as a "memory refresher" and would allow a reader to better recall the information, even if this remains at "gist" level.

Re-reading, despite incurring additional "time costs", can be seen as a beneficial strategy when the level of encoding required by the text is high, again supporting our prediction of higher rates of re-reading, for premises and older readers in particular. This is further supported by the finding that it takes longer for older people to generate a mental model (Gilbert et al., 2004), but once this has taken place, the making of inferences is unaffected. Additional evidence for general slowing in reasoning but not reasoning accuracy , is shown through older and younger readers having similar end point model comprehensions but different reading times and re-reading likelihoods (Stine-Morrow et al., 2004).

In Chapter 2 we established, not only that spatial models are generally harder to comprehend than non-spatial ones. While this may not be completely accurate (see Chapter 3), it is known that eye-movements are influenced by the text which is being processed (Rawson, et.al., 2000; Margolin & Snyder, 2017; Wong & Moss, 2021). Furthermore, regressions are more frequent in texts which are harder to interpret (Inhoff, et.al., 2017). Therefore, by investigating the reading time, regression frequency (re-reading) of the text as a whole as well as premise words this eye-tracking study allows us to investigate if spatial situations are “harder” than non-spatial ones. If spatial texts are truly harder to process than non-spatial ones we predict an increased reading time as well as an increase in re-reading.

This Study

Chapter 4 had three main aims. Firstly, we wanted to deepen our understanding of situational ambiguity through the use of eye-tracking. The use of eye-tracking allows us to investigate “premise re-reading”, in short we can investigate if words/lines containing vital information for the correct construction of a situational model are re-read more often or read for longer. This allowed us to investigate the supposed difference in model comprehension between spatial and non-spatial situations that were found in Chapter 2, where our findings may have been influenced through the introduction of the “4th actor/object”. Therefore it was possible to investigate the reading/rereading rates for ambiguous, unambiguous, spatial and non-spatial stimuli. Furthermore, it was possible to analyse how long words of the premise were read in comparison to other words.

Secondly, we wanted to briefly investigate the “risky reading hypothesis” (Rayner et al., 2006). We investigate only specific aspects of “risky reading” such as word skipping and the re-reading of words but omit some important aspects of the risky reading hypothesis. This is because the stimuli used for our investigation into situational models are not sufficiently similar in terms of word lengths, readability (Kincaid, 1975) and number of words per line. We therefore omit some aspects of “risky reading” such as saccade length (for both inter and intra line saccades) but do investigate fixation durations and number of fixations which are useful for understanding the reading of described situations. As such our findings may not be a true reflection of “risky reading”.

It may be difficult to differentiate risky reading and any “adopted” reading behaviours arising from described situations and ambiguity. It is likely that older readers are more likely to make regressions and more likely to re-read the text, (Rayner et.al.,2007;McGowan & Reichle, 2018). But it is not possible to tell these apart from re-reading as a means to “better” encode the information.

It is currently unknown how, if at all, younger and older readers differ in terms of eye-movements when reading described situations. It is possible that older readers process the two premises differently to those of younger readers and break the text down into smaller chunks (Swets et.al., 2007), likely due to a reduced working memory capacity (Bopp & Verhaeghen, 2007& Borella et al., 2008. Which, in turn, results in an increased reading time and more premise specific regressions.

This study "remedies" the issue of the introduction of the "4th Actor" in questions about spatial situations by asking "Which object/subject was between the two others" in order to equalise spatial and non-spatial questions. The difference between the "4th actor" style questions and "what's in the middle" is investigated in far greater depth in Chapter 6. Both older and younger participants read a set of 48 single paragraph stimuli while we tracked their eye-movements. With the use of eye-tracking testing older readers, and measuring reading time more accurately it is possible hypothesise:

-Hypothesis 1.1: Similarly to our findings in Chapter 2, ambiguous described situations will lead to significantly higher question reading/reaction time than non-ambiguous described situations.

-Hypothesis 1.2: Similarly to our findings in Chapter 2, ambiguity detection accuracy will be lower than model comprehension accuracy, but due to the correction of the "4th actor" spatial comprehension will no longer be significantly lower than non-spatial comprehension.

-Hypothesis 2.1: Older readers will be significantly more likely to skip words in the text as a whole than younger readers.

-Hypothesis 2.2: Older readers will be significantly more likely to regress (move back in the text) while reading compared to younger readers.

-Hypothesis 3.1: "Premise words" will be re-read significantly more often than "non-premise words".

-Hypothesis 3.2: “Premise words” will be read (fixated) for significantly longer than “non-premise words”.

-Hypothesis 3.3: “Premise words” will be significantly less likely to be skipped by readers than “non-premise words”

Method

Participants

For the older participants’ group , we collected data from 20 native English speakers aged 60+ years (Mean age= 70.23 years SD=6.2 years; 16 female). For the younger participants’ group, we collected data from 20 undergraduate students at Bournemouth university, (Mean age= 21.42 years;, SD=4.8 years; 17 Female) who were native English speakers. All participants were recruited through opportunity sampling and all were naïve as to the ambiguity manipulation. The study was approved by the Bournemouth University Research Ethics Committee (ID 27563). Each participant was informed of the experimental procedure and provided informed consent. Due to potential cultural biases (Andrews et al.,2013; Maass et al., 2007) the study was limited to native English speakers living in the UK. Participants received monetary compensation (£10 Marks and Spencers vouchers) or university course credit (1 credit) regardless of how they performed in the study.

In order to ensure that data was only collected from older participants not experiencing atypical ageing (e.g., Alzheimer’s disease), participants were pre-screened using a paper version of the Montreal cognitive assessment (MoCA). Express consent for this was also given.

Experiment Design

Each participant was shown a set of 48 stimuli, presented in pseudorandom order : 24 describing a spatial situation and 24 describing a non-spatial one (see Figure 1a for an example stimulus) . Twenty-four of these paragraphs described an ambiguous situation, while the other 24 described an unambiguous situation, the trial order was determined through a Latin Square counterbalance.

The experiment therefore had a 2x2x2 within-item design, non-Spatial vs spatial, ambiguous vs unambiguous and Old vs Young participants. Each stimulus paragraph was followed by two questions. The first question probed the model and was the same regardless of ambiguity as it probed the “centre” of the model, i.e “Who is neither the fastest nor the slowest runner?” rather than “Who is the fastest runner?” (see Figure 1b for an example & Chapter 5 for further information). The questions had a multiple choice of four answers where; two from viable but indeterminate models for ambiguous stimuli of which one was the correct answer for the unambiguous situation, one option “ There is not enough information” which was the “correct” answer for ambiguous situations and one option which was always incorrect (also not viable in indeterminate situations. The second question probed an aspect of the stimulus, unrelated to the described situation, and was used to test general comprehension of the text (In the same style as found in chapter 2, however 24 additional questions were added to ensure one question for each stimulus paragraph) . These questions were also multiple choice with four selectable answers.

Figure 1a:

Unambiguous non-spatial stimulus (premises highlighted in yellow)

The old chapel relied on donations from the local community. The Alder family had donated a great deal of money to the old chapel, but it was less than the amount of money that the Dwight family had managed to raise for the chapel. However, Mr. Roberts, a new property developer that had just arrived, wanted to grease the wheels of his business plans. He had been spending generously in the village. When Mr. Roberts donated to the chapel, he donated less than the Alder Family.

Figure 1b:

Example of a model comprehension question

Who has donated neither the most, nor the least amount of money?

Descriptive information about the stimuli

The passages in this study were almost identical to those shown in chapter 2, however some changes for readability and flow were made but the premises were not changed. Readability analysis used the readability package for R with the Flesh-Kincaid measure (Kincaid, 1975) using the “quanteda” package (Benoit K, et.al 2018), while the word frequency analysis used SUBTLEX-UK (Keuleers, Lacey, Rastle and Brysbaert, 2011). Table 1 shows the mean readability statistics for the stimuli and questions.

Table 1:

Means table showing readability metrics

Condition	**Ambiguity**	**Word count**	**Word Length in letters**	**Readability**	
				Flesch-Kincaid	**Word Frequency**
			Stimuli		
Spatial	Yes	98.61 (6.41)	4.52 (0.24)	8.05 (1.42)	6.05 (0.13)
Spatial	No	98.96 (5.9)	4.56 (0.24)	7.99(1.45)	6.04 (0.13)
non-Spatial	Yes	94.07 (6.41)	4.52 (0.24)	8.18 (1.39)	6.06 (0.14
non-spatial	No	93.42 (8.31)	4.56 (0.24)	8.22 (1.41)	6.04 (0.13)
			Questions		
Spatial	Same	22.22 (5.65)	4.17 (2.17)	7.34 (2.28)	6.15 (1.17)
non-Spatial	Same	18.82 (4.34)	4.38 (2.25)	7.79 (1.71)	6.10 (1.23)

The Ambiguous paragraphs were constructed to be almost identical to the non-ambiguous paragraphs except that one premise was altered to make the situation ambiguous (e.g "left of" was changed to "right of").

The paragraphs used all possible three object/person relationships, of the 8 ways a 3-object situational model can be arranged 4 are ambiguous and 4 are non-ambiguous which are represented in the stimuli three times per relationship. The premises were un-nested and presented in separate sentences within the paragraphs (for a full list of stimuli and the premises within them please refer to the table called "Stimuli_Prem" in the Stimuli folder for this thesis). Word skipping, for the analysis regarding risky reading, was determined through words which were not fixated but had fixated words before and after.

Apparatus

Eye-movements were recorded with an SR Research EyeLink 1000 eye-tracker at 1000 Hz using a tower-mount set-up. Viewing was binocular, but participants were asked for their dominant eye and only this eye was tracked. If participants didn't know their dominant eye, the Porta test (Roth, 2002) was used to determine it. The Participants' head was stabilised with a chin-and-forehead rest. The text was presented on a Cambridge research systems LCD ++ monitor (resolution: 1920 x 1080 pixels; refresh rate: 120 Hz). The text was formatted in a monospaced Consolas font and appeared as left-aligned black letters over white background. The stimuli were centred vertically and appeared with a X=150 and Y=100 offset with double-spaced lines. The distance between participants' eyes and the monitor was 80 cm. The experiment was programmed in Matlab R2014a (MathWorks, 2014) using the Psychtoolbox v.3.0.11 (Brainard, 1997, Pelli, 1997) . The experiment was run on a Windows 7 PC.

Procedure

. Participants gave their informed consent at the start after having been verbally run through, the MoCA (older readers only) and the study as well as having the run-down on paper to keep. All participants were told to answer “There is not enough information” if there was not enough information for a determinate answer (i.e an Ambiguous situation) and were informed that they would be reading about non-spatial and spatial situations. A 9-point calibration was performed before the experiment. Calibration accuracy was monitored with a drift check before each trial. Recalibration occurred whenever the error was >0.5°. Each trial started with a black gaze-box centred at the first letter of the passage. Once the gaze-box was fixed, it disappeared, and the text was presented on screen.

Participants clicked the left button of the mouse to indicate they had finished reading the paragraph. Each text was followed by two questions, the first question probed the situational model while the second probed some other aspect of the text. The second question was used as a comprehension/attention check question. The experiment lasted about 25–60 mins and participants could take a break whenever they needed.

Analysis

The data was analysed using (Generalised) Linear Mixed Models using the lme4 package v.1.1-21 (Bates et al., 2015) in the R software v. 4.0.4 (R Core Team, 2021). We used dummy contrast coding for Ambiguity (Ambiguous 1, Unambiguous 0), stimulus type (ST) (Spatial 1, non-Spatial 0) and Age (Young 1 , Old 0) . Millisecond reading times were log transformed for the question reading times. Random effects of participant and item were included unless these caused the model to not converge. The full reproducible code is available in Supplementary materials and the OSF (OSF) .

Results

Exclusion

The Montreal cognitive assessment (MoCA), (Nasreddine et al., 2005) was administered to older readers.Procedure dedicated that older readers that scored below 26 points in the MoCA test would not be able to participate in the study.The MoCA was not administered to younger readers, but older readers were informed of the purpose of the MoCA. Participants which scored below 69% in comprehension questions were excluded from data analysis, this removed three younger participants. We excluded trials which were accidentally “skipped”, through accidental mouse button presses during stimuli reading, this process removed 21 individual trials from older readers and 12 from younger readers. We also removed trials where eye-tracking was “unusable” with fewer than 50 discernable word related fixations. This process removed 58 individual trials from older readers and 7 from younger readers. We did not exclude entire participants due to the latter two constraints as the data which remained was still usable. In the case of one individual only 4 trials remained usable due to poor eye-tracking, however their model and general comprehension remained high. In total 37 participants and 82% of the data entered the final analyses.

Question accuracy

As in prior Chapters we investigated question accuracy by condition. The GLMM which replicates the effects of prior studies is located in the supplementary materials of this thesis. Table 2.0 shows results in line with the findings of our prior research. Participants had lower accuracy for questions about spatial stimuli than non-spatial ones. Additionally, ambiguous stimuli cause lower comprehension accuracy (i.e ambiguity detection) than unambiguous stimuli (i.e model comprehension). Further, older readers do not differ significantly in terms of accuracy from younger ones. As shown in Table 2.1 we find no significant differences between any of our experimental conditions despite these being strong interactions in the previous experiment. It may be that there are significant differences as to how in person eye-tracking and online behavioural data is created with participants in the lab performing better due to more direct guidance at the start of the experiment. Additionally the low number of participants in this study may make it difficult to gain sufficient power to detect previously (In Chapter 2) found effects. Alternatively, due to the removal of the “4th actor” in the questions we may have resolved the ambiguity deficits as well as the spatial vs non spatial processing problems. For the time being we will discount these findings until we know more.

Table 2.0:

Mean question accuracy by condition (SD in parentheses)

		Young	Old
Ambiguity	Subject Type	Mean accuracy %	Mean accuracy %
Ambiguous	Spatial	0.208 (0.406)	0.212 (0.408)
Ambiguous	non-Spatial	0.295 (0.456)	0.312 (0.463)
Unambiguous	Spatial	0.656 (0.475)	0.682 (0.466)
Unambiguous	non-Spatial	0.635 (0.482)	0.737 (0.441)

Table 2.1:
GLMM Results for question accuracy as a function of ST,Ambiguity and AgeGroup.

Fixed effects	Question accuracy				
	b	**SE**	**df**	**z value**	**Pr(>|t|)**
Intercept	0.544	0.4876		1.116	.264
Spatial vs Non-spatial	-0.6382	0.6678		-0.956	.339
Ambiguous vs Unambiguous	-0.6518	0.6876		-0.948	.343
Old vs Young	0.4235	0.3137		1.35	.177
Spatial*Ambiguous	1.5939	0.9624		1.656	.097
Spatial*Old	0.0293	0.3211		0.091	.927
Spatial*Ambiguous *Old	-0.6948	0.5009		-1.387	.165

Question reading time

We investigated question reading time. Table 3 shows the mean question reading time by condition between older and younger readers.Table 4 shows the results of a LMM that investigates question reading time as a function of ST, Ambiguity and AgeGroup.The results of the LMM are consistent with earlier findings (Kemper et al.,1993; Liu et al., 2017), that older readers read the questions significantly slower than younger readers. Further, as in Chapter 2 we found that spatial questions were read for a significantly longer time than non-spatial ones. However, in contrast to Chapter 2 we found that there is no significant difference in QRT between ambiguous and unambiguous stimuli, nor is the interaction between ambiguity and paragraph subject. This indicates that ambiguity may not significantly influence processing time during question reasoning. Lastly the lack of significant difference between ambiguous spatial and unambiguous spatial questions, both of which take longer than non-spatial questions, imply that the confound of the “4th actor/object” may not have been a major reason for reader’s difficulty in answering spatial questions.

Table 3:
Mean question reading time (ms) by condition (SD in parentheses)

Ambiguity	Subject Type	Young Mean reading time (ms)	Old Mean reading time (ms)
Ambiguous	Spatial	7060 (4098)	9480 (7102)
Ambiguous	non-Spatial	4950 (2550)	6680 (4379)
Unambiguous	Spatial	6420 (5188)	8460 (7342)
Unambiguous	non-Spatial	5130 (2556)	6290 (3637)

Table 4:
Results of a linear mixed model of Question reading time as function of ST,Ambiguity and AgeGroup.

Fixed effects			**QRT log(ms)**		
	b	**SE**	**df**	**t value**	**Pr(>|t|)**
Intercept	8.639	0.068	79.764	126.878	**<.001**
Younger readers vs Older readers	-0.244	0.082	49.248	-2.957	**.004**
Spatial vs Non-Spatial	0.285	0.075	48.994	3.767	**<.001**
Ambiguous vs Unambiguous	-0.027	0.075	48.994	-0.36	.721
Younger readers Spatial vs Older readers spatial	0.031	0.053	3465	0.581	.562
Younger readers Ambiguous vs Older Ambiguous spatial	0.072	0.053	3465	1.345	.178
Ambiguous Spatial vs Ambiguous non-spatial	-0.131	0.107	48.994	-1.226	.225
Younger readers Ambiguous Spatial vs Older readers Ambiguous Spatial spatial	-0.044	0.075	3465	-0.589	.555

Risky reading

Skip rate

In order to replicate behaviours described in the risky reading hypothesis we investigated word skipping frequency between older and younger readers. Table 5 shows the mean skipping frequency of older and younger readers. Furthermore, we investigated skipping frequency through a GLMM (Table 6), which investigated the frequency of skipping words as a function of Ambiguity, premise words (PW in table, premise word 1, non-premise word 0) and Age group. Premse words were determined to be the words which are required to accurately create a situational model from the text. Skipping was significantly more prevalent in older readers than in younger readers. Additionally, premise words were skipped significantly less often than non-premise words but premise words were skipped significantly more often by younger readers than older readers. However, when the situation was ambiguous, younger readers skipped these premise words significantly less often (*Mean=0.216, SD=0.411*) than older readers (*Mean=0.333, SD=0.471*). Lastly we found that both older and younger readers skipped premise words significantly less often in ambiguous situations.

Table 6:

Results of a linear mixed model of Skip frequency as function of Ambiguity, Premise word (PW) and AgeGroup (Age).

Fixed effects			**Skip**		
	b	**SE**	**df**	**t value**	**Pr(>|t|)**
(Intercept)	-0.788	0.114	-	-6.914	**<.001**
Age	-0.873	0.186	-	-4.677	**<.001**
PW	-0.135	0.045	-	-2.96	**.003**
Ambiguity	-0.092	0.060	-	-1.54	.123
Age*PW	0.202	0.084	-	2.392	**.016**
Age*Ambiguity	0.268	0.039	-	6.82	**<.001**
PW*Ambiguity	0.370	0.065	-	5.695	**<.001**
Age*PW*Ambig	-0.278	0.117	-	-2.379	**0.017**

Regression probability

We investigated a participant's likelihood to re-read a previously read word as a function of Age Group (Old 0, Young 1), ST (non-spatial 0, spatial 1), Ambiguity (Ambiguous 0, Unambiguous 1) and whether or not the word was a premise word (No 0, Yes 1). The results of the GLMM are shown in Table 7. We found that older readers are significantly more likely to regress to a previous word than younger readers but that both older and younger readers were significantly more likely to make more regressions to premise words. Further, we found that when reading unambiguous stimuli younger readers regressed more than older readers, potentially indicating a different encoding strategy. Furthermore, we found that spatial premise words were regressed to less often than non-spatial premise words.

Table 7:
GLMM results table for regression likelihood as a function of Age,Stimulus type (Spatial/non-spatial) (ST), Ambiguity and Premise Word (PW).

Fixed effects	**Regression likelihood**				
	b	**SE**	**df**	**t value**	**Pr(>|t|)**
Intercept	0.608	0.115	-	5.263	**<.001**
Age	-0.540	0.184	-	-2.933	**.003**
PW	0.307	0.058	-	5.264	**<.001**
Ambiguity	-0.132	0.072	-	-1.833	.066
ST	0.0014	0.071	-	0.021	.983
Age*PW	-0.162	0.091	-	-1.767	.077
Age*Ambiguity	0.100	0.043	-	2.321	**.020**
PW*Ambiguity	-0.033	0.083	-	-0.402	.687
Age*ST	0.072	0.041	-	1.757	.0789
PW*ST	-0.293	0.081	-	-3.622	**<.001**
Ambiguity*ST	-0.019	0.101	-	-0.195	.845
Age*PW*Ambiguity	-0.028	0.131	-	-0.213	.831
Age*PW*ST	0.164	0.129	-	1.272	.203
Age*Ambiguity*ST	0.007	0.059	-	0.125	.900
PW*Ambiguity*ST	0.244	0.116	-	2.087	**.036**
Age*PW*Ambiguity*ST	-0.115	0.185	-	-0.624	.532

Fixation Duration and Number of Fixations

We investigated the number of fixations and the duration of these fixations, as a factor of age and whether or not the fixated word was a premise word. These are shown in Table 8 below. We found that premise words were fixated for significantly longer than non-premise words and that, while age group was not significant on its own, younger readers fixated premise words for significantly less time than older readers. Younger readers made significantly more fixations than older readers and fixated premise words significantly more often. The latter being related to the higher regression likelihood found in the analysis shown in Table 8..

Table 8:

LMM results showing fixation duration (top) and number of fixations (bottom) as a factor of Age and Premise word

Fixed effects	**Reading time (ms)**				
	b	**SE**	**df**	**t value**	**Pr(>|t|)**
Intercept	208.77	6.763	36.33	30.86	**<.001**
Age	3.55	11.739	34.74	0.303	.764
PW	10.91	1.445	103337	7.55	**<.001**
Age*PW	-10.59	2.374	104442	-4.46	**<.001**

Fixed effects	**Number of fixations**				
	b	**SE**	**df**	**t value**	**Pr(>|t|)**
Intercept	68.3	3.91	44.2	17.495	**<.001**
Age	4.11	6.45	34.9	0.637	**<.001**
PW	8.24	0.644	10500	12.795	**<.001**
Age*PW	1.64	1.06	10400	1.555	.12

Discussion

This study managed to replicate some of the findings of previous research such as Rayner et al., (2006) and Paterson et al., (2013) in that we found that older readers are more inclined to skip words, re-read words and have longer saccades. However,we did not find any significant differences in fixation duration between younger and older readers, and only found this effect for the premise words. We cannot explain why the effects of fixation duration outlined in Paterson et al., (2013) were not replicated here.It may be that because of the nature of the experiment, both old and young had to pay close attention to the words in the text in order to answer the comprehension question and therefore fixate on them longer. However, it is very difficult to know whether or not the re-reading of older readers comes as a result of risky reading or as a result of the nature of the experiment, as younger people also re-read premise words more often.

We found no significant differences between older and younger readers in terms of accuracy. It is likely that as prior research suggests that older and younger readers do not differ in their ability to create a mental model (Radvansky et al., 1990; Copeland & Radvansky, 2007). However, we also find no evidence of difficulty in processing spatial (ambiguous or unambiguous) situations compared to non-spatial ones and no difficulty in detecting ambiguity.This may be due to the lab based nature of this study where, younger participants were university students (which should have a high reading level) and older participants were members of the ADRC (Aging Dementia Research Center, 2023) and the U3A (University of the Third Age, 2023), highly motivated and educated. As such these participants may have had significantly higher reading ability than those recruited online. Furthermore, participants were verbally briefed as to the nature of the experiment and may have performed better in terms of ambiguity detection because of this. Lastly the number of participants is significantly lower than that of the previous study (Chapter 2) which may indicate that this study does not have sufficient power for the previously found ambiguity detection and spatial effects to be detected. We therefore refrain from making inferences about question comprehension and focus on the eye-movement data.

We also found that older readers were more inclined to skip words. Younger readers made more fixations than older readers and their saccades were significantly shorter which is reflective of fewer words skipped. These indicate that younger readers have a tendency to make shorter saccades with more fixations than older readers, indicating that older reader's saccades are skipping over more content with their longer saccades. This is also consistent with prior findings of risky reading (Rayner et al., 2006) whereby older readers are more likely to skip words than younger readers. Furthermore, while not significant, younger readers' fixations are slightly shorter than those of older readers, which indicates that older readers may have some difficulty in gleaning the meaning of a word and/or encoding the information to memory and may come as a result in "packaging" the text into smaller, more manageable chunks of information (Swets et al., 2007). Alternatively, the longer fixations may indicate that older participants found it harder to combine the two premises (Copeland & Radvansky, 2007), while also having a harder time in filtering irrelevant (to the construction of a mental model) information (Hasher & Zacks, 1988).These findings are in line with the risky reading hypothesis and our hypothesis 2.1. Additionally we found that older readers took significantly longer to read/answer the questions, indicating that while older readers are more inclined to skip more words, they also read for longer in general, consistent with prior research (Myerson et al., 1990; Rayner et al., 2006). Additionally we found that older readers are significantly more likely to re-read previous words, which may also increase their reading time. These findings are consistent with risky reading and re-reading is consistent with Hypothesis 2.2.

In terms of re-reading we found that both older and younger readers were significantly more likely to re-read premise words, such as "better" or "left" , which are vital information for understanding the situational model described. This finding is consistent with hypothesis 3.1. It is difficult to tell this apart from typical “risky reading” behaviour in older readers Rayner et al., 2006. However,it is likely that re-reading in both old and young is used to better enshrine the premises into memory (Margolin & Snyder, 2017; Rawson, et.al ,.2000; Stine-Morrow et.al ,. 2004), potentially indicating that the premises are not integrated during the reading of the text but rather recalled when needed for the question. In either case re-reading of the premise words is likely a boon to both older and younger readers. Potential future research should make use of “neutral” texts, which do not contain premises, in order to compare the “risky” reading behaviour of older readers to the re-reading of important information.

Both older and younger readers are more inclined to re-read premise words and are significantly less likely to skip premise words, indicating that both older and younger readers are aware of what the premises are and the need to look for them. This behaviour is consistent with prior findings that readers in general remember the location of important information and can return to it later as a “memory refresher” (Inhoff, et.al., 2019). Speculatively, it may be the case that older readers shift their primary “risky reading” strategy into “overdrive” when complex text based memory tasks are required of them. It is already known that older people, or at least people with lower working memory capacities break down chunks into smaller pieces (Swets et al., 2007) and that older people exhibit compensatory reasoning mechanisms for low working memory capacity (Piefke et al., 2012). Additionally, prior research indicates that older readers should have comprehension difficulty (Van der Linden et al., 1999; Schurer et al., 2020), but this is not exhibited in this study, which may indicate a “compensatory” strategy employed by older readers. If this is the case then such a strategy, while taking more time, is a boon to older readers. It is also known that when reasoning with mental models, a reasoner is likely to adopt “their own” strategy (Schaeken et al., 2000) and so the adaptation of compensatory strategies may not be too far fetched.

Older readers remain on par with younger readers in terms of comprehension. In combination with the higher the higher skip rate of older readers it is possible to suggest that while premise words are skipped more often by older readers in the first pass, they are then significantly more likely to regress back to these skipped words at a later time. Not only is the return to previous research e.g (Inhoff, et.al., 2019) but also is an important element of “risky reading” (Rayner et al., 2006), Furthermore, these findings are consistent with hypotheses 3.1 & 3.3.

In relation to Hypothesis 1*“Ambiguously described situations have significantly higher question reading/reaction time than non-ambiguous described situations”* we fail to support this hypothesis as our findings did not show any significant differences in the reading/reading time of questions based on whether or not these were ambiguous or unambiguous. The “non” difference between the reading time of ambiguous and unambiguous questions indicates that our initial assumption from Chapter 2 was incorrect and that readers do not create a mental model during questioning but do so during the reading of the situation. We also found that premise words are fixated for significantly longer than non-premise words. Not only is this in line with hypothesis 3.2 but also signifies that those premise words play a significant role in encoding situation information. Although, ambiguity may only be detected when participants are probed about the model, as even if a situation has multiple solutions at least one of the objects/subjects are in a position which is unambiguous. However, this study differs in terms of experimental setup from Chapter 2 which was completely online.

While it is not yet known if encoding strategy differs between ambiguous and unambiguous, non-spatial and spatial texts, the comprehension results from Chapters 2 could indicate that this is a potential explanation. Further, evidence for situation models being created during stimulus reading come from the increased re-reading/regression rates for younger people reading ambiguous stimuli. Indicating that there is some difficulty in integrating two premises, if these premises describe a situation that is ambiguous.

We found that spatial premise words were re-read significantly less often than non-spatial premise words. However,this could be explained through the fact that non-spatial premises often contained names of people which may/ or may not be harder/easier to memorise, such investigation however is beyond the scope of this study and this thesis. However, as we speculated in Chapter 2, the creation of a mental model from a spatial situation may be influenced by semantics (Knauff & Johnson-Laird 2002;Knauf & May, 2006) compared to non-spatial situations. We also found that there are no significant differences in the re-reading of spatial and non-spatial premise words between older and younger participants, which, assuming mental models are created during the reading of the situation, holds true with prior findings that older and younger people do not differ in their ability to create a mental model (Radvansky et al., 1990). Lastly, we found a significant interaction between premise word re-reading in ambiguous spatial situations, this is, perhaps, a more likely indicator of different encoding strategy.

In short the findings of this study indicate that both older and younger readers are both capable builders of mental models from situations described in text and, in line with prior findings (e.g., Radvansky et al., 1990;Gilbert et al., 2004) do not differ in their ability to create and reason with a mental model. However, we suggest that the strategy in which older and younger readers use to do so may differ, likely as older readers use some as of yet unclear compensatory mechanisms (Piefke et al., 2012), such as frequent re-reading (Margolin & Snyder, 2017; Rawson, et.al ,.2000; Stine-Morrow et.al ,. 2004) or the breaking down of text into smaller “chunks” (Swets et al., 2007). We also found strong evidence for “risky reading” in older participants but not younger ones, further indicating that older readers adopt certain compensatory strategies for lower working memory capacity.

Conclusion

In conclusion we managed to replicate the risky reading hypothesis in that older readers skip more than younger readers and that they re-read more. However, we are unable to differentiate re-reading of premises in older readers from “risky reading” (Rayner et al., 2006). With that said we do see some evidence to suggest that different encoding strategies are used between young and old but that both groups re-read premises more often than non-premises as this is likely beneficial to encoding the premises for later retrieval (Inhoff, et.al., 2019). Furthermore, we suggest that there may be some unknown strategies employed by older readers as compensation for working memory deficits (Piefke et al., 2012). We further suggest that, in accordance with prior research (e.g., Dror et al 2005; Radvansky, 1990) that older and younger readers do not differ significantly in their ability to create and make inferences from a mental model as well as there being no significant differences between older and younger participants’ ability to detect ambiguity.

The work carried out up to this point has allowed us to establish a few important facts about situational ambiguity and ageing. Thus far we have reasons to consider the following:

1. Ambiguous situational models do not appear to do not differ significantly in reading time from unambiguous ones, though where a reader’s attention is drawn to does.
2. Non-Spatial situational models are significantly easier to interpret than Spatial-situational models, though they appear to be read in a similar way.
3. Older readers are significantly slower readers than younger readers and read in a significantly more “risky” way than younger readers. However, this may not cause significant differences in comprehension when comparing older and younger readers.
4. Re-reading is likely significantly increased in areas that require special attention to comprehend a situational model.
5. Older readers are significantly more likely to re-read than younger readers.

With this knowledge in hand it is possible to start to gain a rudimentary understanding as to how situational models are processed by readers. Firstly it is likely that re-reading of vital information plays a key role in a reader's understanding of the text. In chapter 4 we found significant increases in re-reading of the premises which make up the situational model. Re-reading, appears to be greatly beneficial for readers to move beyond a "gist" (Sachs, 1967) level memory about a situational relationship. By re-reading we are able to securely encode the vital information in the text (Wang, Bolin, Lu and Carr, 2018).

However, in the last chapter we didn't address the issue of the "4th actor" problem from chapter 2, we only amended it. As indicated by the mean comprehension accuracy it may be the case that even without the introduction of a 4th actor during the question that spatial situations are harder to interpret than non-spatial ones. In the next chapter we will investigate this further to determine whether or not spatial questions with and without 4th actors are actually interpreted all that differently. The confound introduced by accident may actually have opened up an additional question as to whether or not described situations are interpreted differently depending on the question. Moreover this too highlights another as of yet unanswered question, why are some models preferred over others?

When dealing with ambiguous models many reasoners end up preferring one interpretation over others (Goodwin & Johnson-Laird, 2005;Nejasmic, Bucher & Knauff, 2015). As mentioned previously there are many potential reasons for this e.g. syntactic/semantic. There is some debate as to how the preferred meaning in global syntactic ambiguity is determined (Tabor et.al 1997; van Gompel et.al 2000; Green & Mitchell 2006; Clifton & Staub 2008; Vosse & Kempen, 2008) which is beyond the scope of the current research. With that said it is important to note that there is some disagreement as to how these competing syntactic interpretations are processed during reading and whether or not such competition between syntactic interpretations exists. For example, (Frazier,1987) outlines that a single syntactic solution arising from ambiguity is constructed and evaluated one at a time in a serial manner. Though along this line of thought there are different interpretations as to when the preferred interpretation is accepted, either this is chosen at the beginning (Frazier & Rayner, 1982), or it possible that the simplest is preferred, (Frazier & Fodor, 1978). In both cases if an interpretation contradicts subsequent contextual information, this sort of ambiguity forces readers to re-read and reanalyse, as such if readers do not reanalyse for whatever reason, ambiguity has no additional cognitive cost. As such, assuming situationally ambiguous models cause readers to behave similarly, additional reading time may be prevalent for readers who realise the presence of ambiguity, though this would happen during post reading. However, without other influences readers create models "bit by bit" (Johnson-Laird & Byrne, 1991). It is currently unknown as to which "bit" takes priority when constructing a model based on a described situation. It is also not known whether this "priority" influences the model which ends up being preferred, assuming that ambiguity goes unrecognised by the reader. We attempt to answer these questions in a "subchapter" of Chapter 5.

Alternatively, competing or Parallel models, would predict an increased processing load during reading but not arising from analysis from incongruent context or later realisation of ambiguity, rather due to these interpretations being assessed by parsers all at once in parallel, as seen in (McDonald 1994; Tabor et.al 1997; Clifton & Staub 2008). The increase in processing time for the reader is thought to be down to an increased load on working memory where multiple plausible interpretations are considered at the same time. With that said there is some more recent disagreement as to whether or not syntactic ambiguity is processed serially, without competition between interpretations (Van Gompel et al., 2005) or in parallel, with competing interpretations of syntax (Green & Mitchell , 2006). In either case, the meaning of the text can be interpreted in a different way by the reader to what the writer intended, which is especially true if the writer has purposefully written the text to be ambiguous in some way. However, this does not serve to explain how a question could influence the detection of ambiguity.

Chapter 5: Situational Ambiguity and Age

Abstract

Text processing in ageing has been studied extensively in the past. While it is known that text processing (Myerson et al., 1990), working memory capacity (Bopp & Verhaeghen, 2007; Borella et al., 2008; De Beni et al., 2004; Fiore et al., 2012 ; Park et al., 2002) and the ability to reprocess ambiguous text decline with age (MacDonald et al., 1992,Yoo & Dickey, 2017), less is known about how ageing influences the detection of ambiguity in situational models. Here we present two experiments investigating how different questions probing how ambiguity influences the reader's interpretation of situational models in younger and older adults. We further investigated if these probing questions influence how situational ambiguity is detected. Consistent with prior research we found that older participants take longer to encode and make inferences from a mental model (e.g.,Gilbert et al., 2004) but that there are no age related differences in model comprehension and ambiguity detection. We suggest that the longer reading time of older participants is the result of lower text comprehension, as there are no significant differences in question accuracy between older and younger readers, but rather as a result of encoding or reasoning strategies employed by older readers to compensate for a lower working memory capacity. We also found that older participants re-read more often than younger participants (Rayner et al., 2006). Furthermore, we found that probing the "internal" part of the model significantly improves ambiguity detection and that such a probe is an even greater boon to older participants when detecting ambiguity in spatial situations. We suggest that probing the internal part of the model discourages "lazy" strategies and encourages full model integration which heightens the likelihood to detect ambiguity.

Introduction

Originally, mental models were proposed as a programmatic basis for thinking (Craik, 1943). Mental models are thought to be pivotal to the understanding of discourse whereby processing verbal or written structures (premises) lead to an internal or mental model of the described situation akin to the result of perceiving or imagining the situation. These models are used not only for language processing (Johnson-Laird 1983) but also for spatial reasoning (e.g. Vosgerau, 2006). Mental model theory (Johnson-Laird & Byrne, 1991) stipulates that people make inferences from an array they construct from given information. However, there are a few different ideas as to how exactly a mental model is constructed and used, for example “Preferred Mental models” (Ragni & Knauff, 2013) which builds upon the construction and inference “stages” outlined in Mental Model Theory (Johnson-Laird & Byrne, 1991). In this present study we created paragraphs with embedded premises, to simulate “naturalistic” reading and used Preferred Mental Model Theory (Ragni & Knauff, 2013) as a basis to investigate a reasoner’s ability to “detect” if premises describe more than one model. It is currently unclear as to whether or not readers can accurately detect if premises describe more than one model in paragraph reading, though it is known that readers tend to be “blind” to alternative models (Johnson-Laird 1994).

Among the key properties of mental models is that their structure corresponds to the structure of what they represent (like a visual image), and thus that individual entities are represented just once in a model. If there is more than one viable model, the reader generates alternative models sequentially (Johnson-Laird & Byrne 1991). Johnson-Laird & Byrne describe this process by breaking it down into three phases; comprehension, description, and validation. During model comprehension a reasoner uses their knowledge of how the world works to construct their initial model which is then validated by checking the premises. However, mental model theory assumes that only this initial model is required, even if there are additional adjustments that need to be made, and that the premises may be forgotten (Mani & Johnson-Laird, 1982).

There are different theories as to how the construction of mental models occurs. Preferred Mental Model Theory (PMMT) (Ragni & Knauff, 2013) is slightly different to Mental Model Theory (MMT). Specifically, preferred model theory (Ragni & Knauff, 2013) suggest different terminology and mechanisms for the model construction phases, these being; construction, inspection, and variation. PMMT serves to explain the difficulty people have with complex reasoning which are not well explained by MMT. Though the MMT explains some of the difficulty people have when making inferences from complex mental models, the MMT omits explaining how difficulty arises during the construction of mental models.As with Johson-Laird & Byrne's comprehension phase, the In the model inspection phase, a "generally true" version of the mental model is constructed using the premises. This initial model is then inspected by the reasoner to make preliminary conclusions. During this phase the reasoner is able to find relations not directly stated by the premises (Ragni & Knauff, 2013).

Note, however, that Johnson-Laird & Byrne's description phase assumes the reasoner's preliminary conclusion functions as a description of the model as a whole (Johnson-Laird & Byrne, 1991). Therefore, if there are two equally viable models, the reasoner would have to envision the alternative possibilities by two alternative models

(Johnson-Laird, 1994). However, reasoners generally only construct a single model and overlook other potentially viable ones (Johnson-Laird, 1994), i.e. reasoners prefer to construct one model and "check" it for errors. As "erroneous conclusions will tend to be consistent with the premises rather than inconsistent with them", a model which passes such checks becomes a reasoner's preferred template from which to make further inferences. The present study was not designed to address the theoretical implications of whether PMMT or MMT are more or less accurate and generally refer to both theories "equally".

This study expands upon prior work investigating ambiguity detection in text from Chapters 2 & 4 as well as prior work on mental models (e.g., Ragni et al., 2007). Additionally, as in Chapter 4, we also investigate how ageing changes the ability of readers to detect ambiguity. Furthermore, in Chapter 2 we speculated that the reason for low rates of ambiguity detection may be due to participants making inferences using one preferred mental model. This in turn led to the other viable model being "ignored" and thus ambiguity going undetected. Experiment 2 of this study presented participants with a different style of question "Which object/subject is in the middle", which was designed to "discourage" the creation of only a single model from which to reason.

It is known that participants build up the model "bit by bit" (Johnson-Laird & Byrne, 1991) in the case of two premises the first premise is used as a baseline from where the relationship between objects/subjects described in the second premise can be added. For example:

<u>Example 1:</u>

*Premise 1: "The **apple** is to the right of the **orange**"*

Creates the base relationship between objects:

<u>Orange > Apple</u>

*Premise 2:" The **pear** is to the <u>left</u> of the **orange"***

Which in turn "inserts" the **pear** to the <u>left</u> of the orange:

In the case of Example 1, there is only one viable model which can be created this way. It is important to note that the individual premises are also "stored" in working memory (Goodwin & Johnson-Laird, 2005) as these may be required for later revision of the model (Rauh et al., 2005).However, if we change premise 2 to state:

<u>Example 2:</u>

*Premise 1: "The **apple** is to the right of the **orange**"*

Creates the base relationship between objects:

<u>Orange > Apple</u>

*Premise 2: "The **pear** is to the right of the **orange**"*

We create an ambiguous situation where there are two viable models:

<u>Orange > Pear > Apple</u>

And

<u>Orange > Apple > Pear</u>

The placement of the pear, which is "inserted" after the relationship between the orange and the apple (Johnson-Laird & Byrne, 1991), is dependent on the reader's preference. As we have seen in Chapters 2 & 4 readers are typically unaware that two models exist. However, in Chapter 2 we asked questions such as:

<u>Example 3:</u>

" *Peter's hand is on the Apple, does it need to go left or right to reach the pear?*"

Both potential models outlined in Example 2 are "useful" (Nejasmic et al., 2015), and allow the reader to answer the question, with no revision of the model needed (Rauh et al., 2005).

The fact that either of the two potentially preferred models are useful enough to complete the task is the likely reason for the other model being ignored and therefore ambiguity going undetected. Also it is important to note that any object/subject mentioned twice, i.e in relation to another object/subject can never be in an ambiguous location, in the case of Example 2 all objects are described in relation to the **orange**. From this point on we refer to this object/subject as “the anchor”.

Alternatively and somewhat speculatively, a reader may not even need to use a full model in order to answer the question set out in Example 3. In Example 2 the first premise describes the relationship as the **apple**, where Peter's hand is, as being to the right of the **orange.** Using only this information the reader can then infer that the **pear** must be to the left of the apple. In this case a single premise gives the reader enough information from which to infer the location of the third object/subject in a way which is good enough to complete the task. Moreover, this line of reasoning does not violate either premise. Whether this is the case is not yet known but in either case ambiguity remains undetected. From this point onward we will refer to a complete model e.g (“**Orange > Pear > Apple**”) which infers the location of a third object/subject from a single premise as a “lazy model”.

In Chapter 4 we changed the style of question in order to mitigate the potential increased difficulty in spatial situation comprehension and ambiguity detection. The difficulty caused by the “4th actor” (as reflected in the question style shown in Example 3 and in Chapter 2), may have inadvertently caused an increase in the number of objects/subjects (Goodwin & Johnson-Laird, 2005) in spatial models and may have induced complications related to model revision (Rauh et al., 2005). However, by asking “what’s in the middle” for ambiguous situations we may also have discouraged readers from using a single preferred model or a single premise from which to make inferences. For example consider the ambiguous situation:

Example 4:

Premise 1: “***Thom*** *is richer than* ***Anne”***

Premise 2: “***Alina*** *is poorer than* ***Thom”***

Which creates the two viable models (richest left, poorest right):

“**Thom < Anne < Alina**”

And

“**Thom < Alina < Anne”**

We now ask:

“ Who is neither the richest nor the poorest?”

Assuming we have created either model by “inserting” (Johnson-Laird & Byrne, 1991) **Alina**, who is poorer than **Thom,** from the second premise into the model:

“**Thom < Anne”**

Firstly, if we only use a single premise, premise 1 which has given us our baseline, to infer **Alina’s** position in terms of wealth we could infer that **Alina** is wealthier than **Thom,** because **Anne** is poorer than **Thom.** However, this lazy model directly violates premise 2 which states that **Alina** is poorer than **Thom**, and as such this model is far less likely to be preferred (Jahn et al ., 2007).

Secondly, if we use both premises and “insert” (Johnson-Laird & Byrne, 1991) **Alina** into the model in a way which does not violate either premise we are again left with two viable models. Generally it is known that people prefer to create models with the fewest number of assumptions (Byrne & Johnson-Laird, 1989; Goodwin & Johnson-Laird, 2005). The reader may for instance prefer a model that assumes that **Alina** is directly “adjacent” to **Thom** as could be implied by premise two, or that because mentioned later that **Alina** is adjacent to **Anne**. Research into the construction of mental models has found that people use a variety of strategies for constructing and subsequent deduction of a mental model, (Schaeken et al 2000) and while we can not determine the precise strategy used by participants, it is possible to “discourage” the use of strategies which do not aid in model comprehension or ambiguity detection. By probing the “internal” element of the model (“What’s in the middle”) it may be possible to “discourage” the use of a single premise, while else “encouraging” model revision (Rauh et al., 2005), both of which should increase ambiguity detection.

In Chapter 2 we discovered that spatial situations were significantly harder to reason with and detect ambiguity in, compared to non-spatial ones. It is currently not known how or why this is. It is possible that semantic interference (Knauff & Johnson-Laird 2002;Knauff & May, 2006) also plays a role in the construction of a preferred model and subsequent ambiguity detection, but this is not certain. Additionally, when recalling the situation during probing, irrelevant descriptives, e.g.,”It was a sunny day” may also interfere with encoding and recall of the premises “The car was left of the bike” (Dewar et al. 2007; Craig and Dewar, 2018), which may not happen with more abstract concepts such as wealth or popularity.However, if spatial situations are more influenced by semantics readers may be “guided” to create a single preferred model in ambiguous situations, or adopt a lazy model in spatial situations. We therefore repeated the investigations into the differences between spatial and non-spatial models and investigated if the different subject types (ST) lead to different models,preferred by readers.

In this study we not only investigated participant’s model comprehension and ambiguity detection but also if participants built preferred models based on a specific premise. We hypothesised that “discouraging” participants from making inferences based on a single premise, “discouraging” the construction of a single preferred model or “encouraging” model revision may aid detection of ambiguity. We did this by asking participants “what’s in the middle?” style questions. As a result this style of questioning may also increase unambiguous model comprehension. From this point onward we will refer to “what’s in the middle?” style questions as “internal questions” and “ ***X*** is ***Y***, where does ***X*** need to go ***Z***?” style questions as “external”. Therefore asking “internal” questions will likely lead to an increase in both ambiguity detection and in unambiguous model comprehension compared to asking “external” ones questions. While previous research has shown that people generally ignore, are blind to, or omit their non-preferred models (e.g., Johnson-Laird, 1994; Goodwin & Johnson-Laird, 2005) and typically create a model which is “useful” for the task (e.g., Nejasmic et al., 2015) it is unknown how the probing of a model may influence ambiguity detection. Therefore, if probing the “internal” part of the model encourages people to not omit non-preferred models, it stands to reason that people can create multiple models if this is “useful” to the task and that more than one model can be stored in working memory to be reasoned with.

We also investigated whether the likelihood to prefer one model over the other is different when asking internal questions compared to external ones. As this is likely to show if readers are likely to create a single model from which to reason, be it a lazy or preferred model. Further, we investigated if older readers have different preferred models and how the reasoning process may change as we age.

Experiment 1

Experiment 1 investigated the potential ageing differences in ambiguity detection in terms of: Question reading time and Question accuracy. In lieu of eye tracking which is not economical on a large scale, we investigated the reading time for informative lines vs. non-informative lines of text, akin to the investigation of reading/re-reading of premise words in Chapter 4. Further, we also investigated preferred models when participants did not detect ambiguity, and compared older and younger reader’s preferences.

Transparency and Openness

The full de-identified data sets for both Experiments 1 and 2, along with the full reproducible, experimental code, stimuli and preprocessing code, analysis code and other notes can be found via the OSF link provided in the author's note. Due to being a more exploratory piece, this study was not pre-registered nor did we make our hypotheses known prior to investigation. In terms of the LMM and GLMM models we used subject and item as random factors by default, we are aware that more complex models can be used.

Participants

We collected data from 106 (Mean age= 26.6 years, SD=9.1 years; 88 Female, 1 other) young, native English speakers (readers) recruited online from Prolific and undergraduate students at Bournemouth university. We also collected data from 50 older, native English speakers (Mean age= 65.9 years, SD= 5.25 years; 36 Female) recruited online from Prolific (Prolific.co). Participants recruited via Prolific were compensated with £7.56 per hour, and students from Bournemouth University were granted 1 course credit for their efforts. All participants were naïve as to the ambiguity manipulation, though were primed to answer "There is not enough information" if there was not enough information (from an ambiguous stimulus) and were informed that they would be reading about non-Spatial and spatial situations. The study was approved by the Bournemouth University Research Ethics Committee (ID 27563). Each participant was informed of the experimental procedure and provided informed consent. Due to potential cultural biases (Andrews et al .,2013; Maass et al ., 2007) the study was limited to native English speakers living in the UK.

Materials and Designs

The stimuli consisted of 48 paragraphs: 24 describing a spatial situation and 24 describing a non-Spatial one. Each of these paragraphs could be seen in one of two versions: ambiguous or unambiguous. The stimuli were adapted from Chapters 2 and 4.

The experiment had a mixed 2x2x2 within-item design, non-Spatial vs spatial, ambiguous vs unambiguous, older vs younger. A power analysis of our pilot study of older readers revealed that an age/accuracy effect was small and 50 participants would be needed to obtain 80% power to find it. The pilot data, analysis, stimuli and questions can be found on the OSF link provided on the title page. Model comprehension was probed using questions about the situation described in the simulus. The initial question probed the model, which could be either ambiguous (correct answer: "There is not enough information") or unambiguous (single viable option) The second question probed general comprehension of the text. Below we illustrate an example stimulus in text form:

"The town of Rockport hosted three major events that always drew in large amounts of crowds. ***The folk music festival was more popular than the Rockport Grand Prix****. Both were important sources of income in the town.* ***The Sailing festival was more popular than the Rockport Grand Prix****, but this didn't matter to the townsfolk. Rockport was one of the most up and coming towns in the area. The yearly events were beginning to draw in crowds from further and further afield."*

The stimulus described an ambiguous non-spatial situation, the premises (**bolded** in the text), created a situation where it was not possible to tell which festival is the most popular. Another example, this time for a spatial stimulus, is shown below:

"Spice Island lay just three miles offshore. The island didn't have a lot to offer in terms of comfort but there were a handful of people that called it home. ***There was a small weather station to the left of the small dock used to load and unload supplies. There was also a supply store to the left of the docks.*** *It was overpriced because everything had to be imported from the mainland. Many islanders didn't mind spending more on supplies if it meant avoiding a long trip to the mainland."*

<u>Note:</u> *In the experiment participants would not see the full text, only one line at a time, an example of this is shown in Figure 1 in the procedure section.*

Descriptive information about the stimuli

Readability analysis used the readability package for R with the Flesh-Kincaid measure (Kincaid, 1975) using the quanteda package (Benoit K, et.al 2018), while the word frequency analysis used SUBTLEX-UK (Keuleers, Lacey, Rastle and Brysbaert, 2011).Table 1 shows the mean readability statistics for the stimuli and questions.

Table 1:
Mean readability statistics for the stimuli and questions (SD in parenthesis).

Stimuli

Condition	**Ambiguity**	**Word count**	**Word Length in letters**	**Readability**	
				Flesch-Kincaid	**Word Frequency**
Spatial	Yes	85.58 (4.07)	4.44 (0.19)	7.05 (1.62)	5.95 (1.32)
Spatial	No	85.79 (3.75)	4.44 (0.19)	7.04 (1.60)	5.96 (1.32)
non-Spatial	Yes	85.50 (3.69)	4.41 (0.26)	7.17 (1.26)	5.99 (1.15)
non-Spatial	No	85.45 (3.63)	4.40 (0.25)	7.18 (1.25)	6.00 (1.15)
			Questions		
Spatial	Yes	22.1 (5.62)	4.17 (2.17)	7.34 (2.28)	6.152 (1.17)
Spatial	No	23.25 (6.21)	4.17 (2.17)	7.34 (2.28)	6.152 (1.17)
non-Spatial	Yes	17.50 (4.46)	4.42 (0.53)	7.79 (1.71)	6.102 (1.23)
non-Spatial	No	18.16 (4.36)	4.43 (0.42)	7.79 (1.71)	6.102 (1.23)

The ambiguous paragraphs were created by altering one premise of the non-ambiguous paragraphs by changing right to left or better to worse for example. The full list of stimuli in their various conditions can be found on the OSF link provided.

Of the 8 ways a 3-object situational model can be arranged, 4 were ambiguous and 4 are non-ambiguous. These arrangements were evenly distributed across the counterbalanced lists of items. The premises were presented in separate sentences. Descriptive antonyms, e.g UP-DOWN, POORER-RICHER, and were used equally often across the 48 items .

Procedure

Participants indicated their age, any diagnosed reading disability, if they had normal, or corrected to normal vision, and if they were native English speakers. Then they were provided details of study participation needed for providing informed consent. The experiment was created using PsychoPY(Pierce et al., 2019).

Participants would then see a screen with the stimulus text masked by the letter “X” (e.g Hutzler et al., 2013) along with the instruction “Please press the down key to begin reading” (see Figure 1). This mask was used in order to facilitate re-reading measures, as participants had to move the unmasked section in order to re-read previous lines. Further, this would allow for some predictions to be made in terms of the risky reading hypothesis. Pressing the down key would show the next line of text and simultaneously remask the prior line. Up key presses reversed this allowing readers to move back up through the text. The stimuli were always presented in 6 lines. The 7th line was a further instructional text reading “please press space to continue” which would advance participants to the question phase.

To account for different size monitors, the stimuli size and position were defined in PsychoPy's "height units which on a 16:10 aspect ratio screen are represented as -8,-5 as the bottom left and 8, 5 as the top right. These height units are relative to the size of the window, not the screen, though the window would automatically enter full screen mode. For detailed information about PyschoPy's height units please see (psychopy.org/general/units.html#units-for-the-window-and-stimuli).

The text would always appear in the vertical centre of the screen, regardless of screen size and the instruction "Press the down key to begin reading" was displayed at the vertical centre between the first line of text and top of the screen. All text was presented centred on the screen (Davenport & Smith, 1965) in white on a black background using the consolas font with a height of 0.03 units. The experiment was hosted on Pavlovia.org.

Figure 1:
Stimulus as presented on screen with the first line of text unmasked

Participants could read at their own pace and could only advance if they pressed the spacebar when line 7 was unmasked. This line by line self-paced reading technique allowed us to calculate reading time for the individual premises, as each line contained a single sentence. The first five lines contained 15 words, while the last line (line 6) contained between 5 and 16 words. Two of the lines contained the premise information (1 each) did not differ significantly from non-premise lines in word count (Premise lines: mean=14.98,SD=2.36; Non-premise lines: mean=15.1, SD= 2.45) or readability (Premise lines Flesh-Kincaid: mean=4.98, SD=2.47; Non-premise lines Flesh Kincaid: mean=5.061, SD=2.54). The number of lines between premise 1 and premise 2 was not accounted for during counterbalancing but did not differ significantly across conditions (mean= 2.4, SD= 1.05).

Two questions followed each stimulus. The first question asked about the spatial or non-Spatial relationship described in the premises and had 4 answer options (see Figure 2a). The question was displayed at X=0, Y=0.3 at 0.05 height units. The options were displayed at the bottom of the screen in white, equidistant, clickable boxes (Width units 0.38, Height units 0.2) at positions along the Y axis at -0.35 units and the X axis at -0.675, -0.225, 0.225, 0.675 units respectively.Two questions followed each stimulus. The first question asked about the spatial or non-Spatial relationship described in the premises and had 4 answer options (see Figure 2a & b). The question was displayed at X=0, Y=0.3 at 0.05 height units. The options were displayed at the bottom of the screen in white, equidistant, clickable boxes (Width units 0.38, Height units 0.2) at positions along the Y axis at -0.35 units and the X axis at -0.675, -0.225, 0.225, 0.675 units respectively.

Figure 2a:

Question screen showing a spatial question

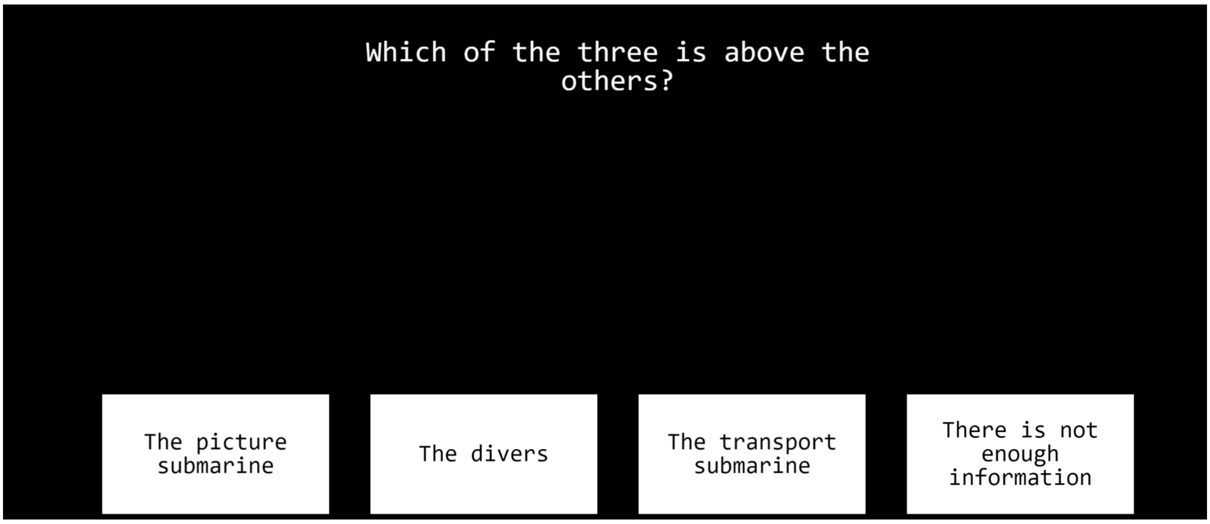
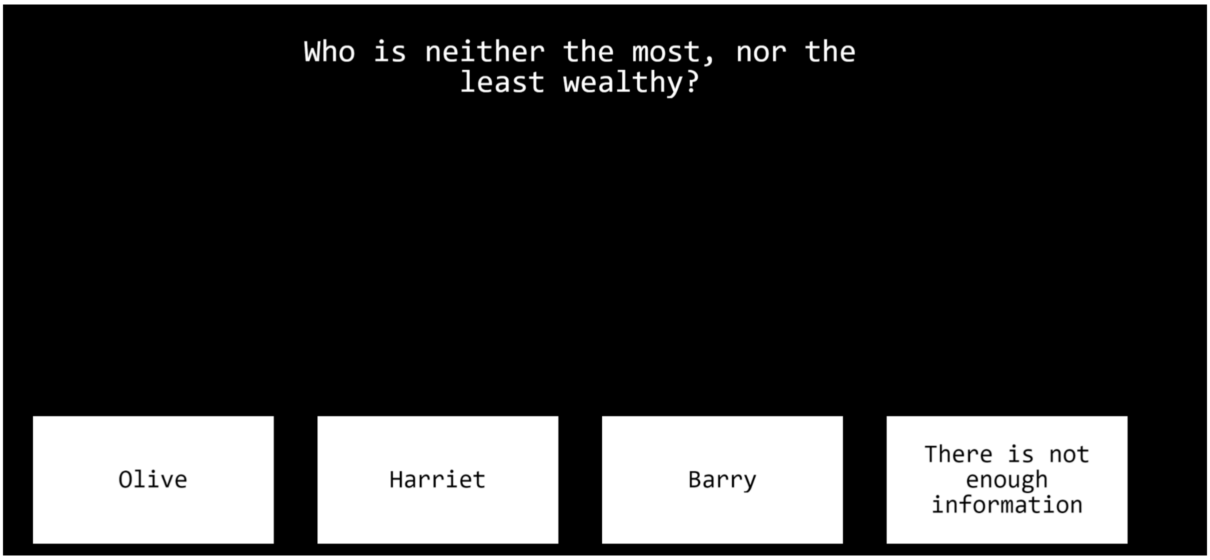


Figure 2b:

Question screen showing a non-spatial Question



These four options always had two options that were equally plausible in the ambiguous condition, though only one of these would be correct in the unambiguous version. A third option was categorically incorrect and a fourth was “There is not enough information” which was always correct for the ambiguous conditions. The questions and answer choices didn’t differ between ambiguous and unambiguous versions of the paragraphs. Participants would indicate their selection with a mouse click. All 48 items were presented in pseudorandom order. Participants completed the study in a single session. The experiment lasted about 25-45 minutes on average and participants could take short breaks when needed. There were no practice items.

Analysis

The data was analysed using (Generalised) Linear Mixed Models using the lme4 package v.1.1-21 (Bates et al., 2015) in the R software v. 4.0.4 (R Core Team, 2021). We used dummy contrast coding for Ambiguity (Ambiguous 1, Unambiguous 0), stimulus type (ST) (Spatial 1, non-Spatial 0) and Age (Young 1 , Old 0) in the analysis. Millisecond reading times were log transformed for both the question and stimulus reading times. Random effects of participant and item were included unless these caused the model to not converge. The full reproducible code is available in Supplementary materials and the OSF (OSF) .

Results

We removed 9 (3 older, 6 younger) participants for low accuracy (below 70%) in general comprehension questions. Additionally, 8 (younger) participants were removed due to having over 18% trial “skip” rates, where their overall reading time of the stimulus was below 5 seconds, this also removed a further 33 individual trials. With these elimination protocols 99 participants (93.01% of the data) remained for analysis. The mean reading times are presented in Table 2.

Reading time of Stimuli

Firstly, we investigated the Stimulus reading time (SRT) in their single paragraph format. The means and standard errors for SRTs are presented in Table 2. It is clear that there are no significant differences in stimulus RT based on spatial or non-spatial situations and that ambiguous and unambiguous stimuli were read for a similar amount of time. However, as expected we found that older readers take significantly longer to read the stimulus than younger readers, this is consistent with prior research and our findings in Chapter 4

Table 2:

Mean stimulus reading times in ms (SD in parenthesis).

	Stimulus RT (ms)	
Stimulus Condition	**Young**	**Old**
non-Spatial unambiguous	38532 (40560.94)	46635 (27339.99)
Spatial unambiguous	43233 (96461.76)	48400 (39037.26)
non-Spatial ambiguous	40354 (41969.53)	48689 (27341.35)[2]
Spatial Ambiguous	41304 (51926.39)	50062 (80985.24)

2

Question reading time

The question reading time (QRT) means and standard deviations are shown in Table 3. We found two significant differences in QRT. Firstly, younger participants read questions significantly faster than older readers, consistent with the findings in Chapter 4 and prior research. We also found both young and old readers read questions about spatial text for longer, this can be seen in Table 3 and Figure 3. This may indicate that questions about spatial situations are harder to reason with than non-spatial ones, again consistent with our findings from Chapter 4. We also found that older readers take significantly longer in general to answer questions compared to younger readers which may indicate difficulty in memory retrieval in order to make inferences about the model.

Table 3:

Mean comprehension and situational question reading times in ms (SD in parenthesis) .

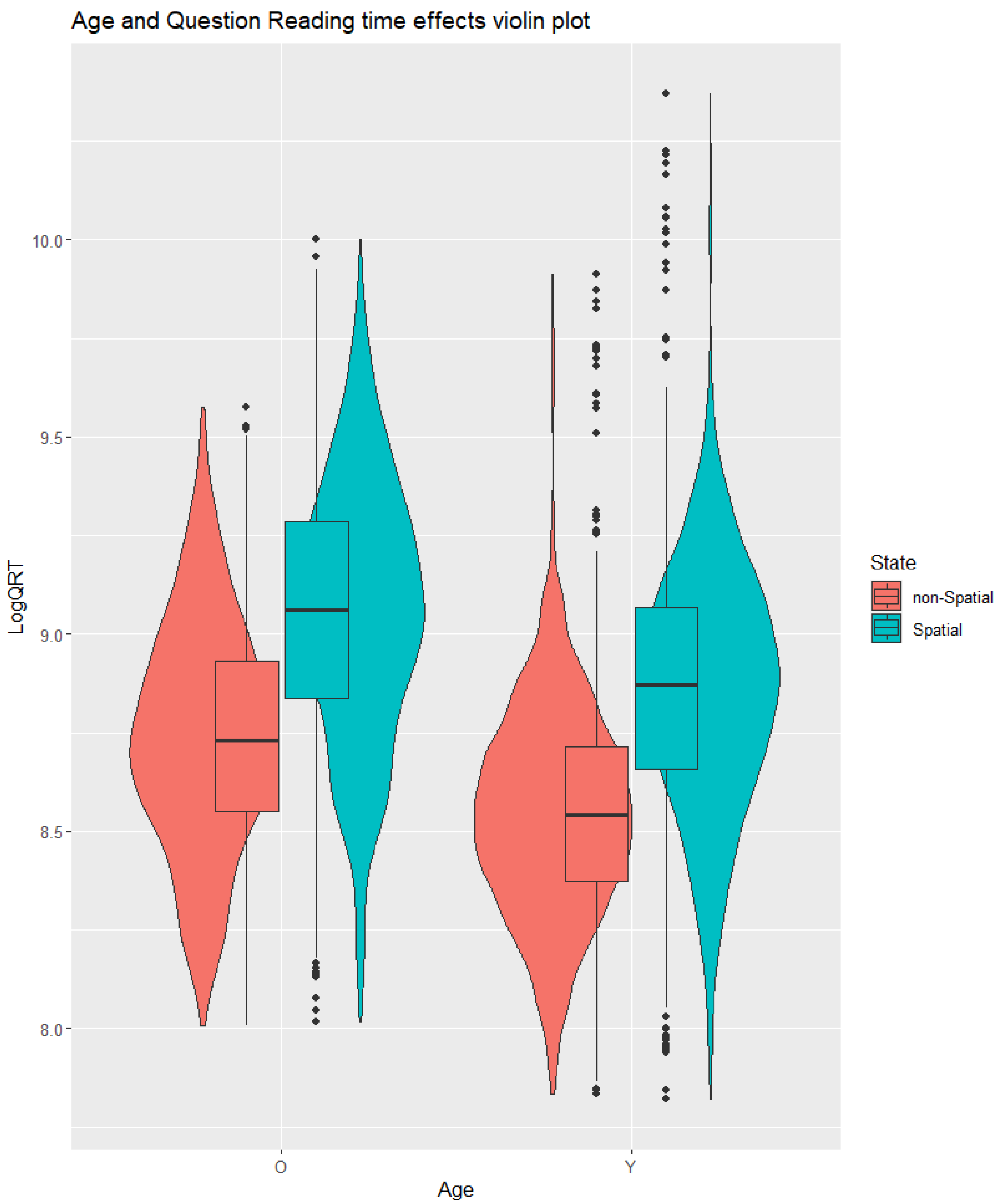
	Question RT (ms)			
	General Comprehension Questions		Situation Model Questions	
Question Condition	**Old**	**Young**	**Old**	**Young**
Non-spatial unambiguous	4490 (3357.14)	4232 (4769.98)	7194 (5281.61)	7021 (13553.16)
Spatial unambiguous	4323 (2594.56)	4802 (9508.40)	11143 (9235.18)	9794 (10516.56)
Non-Spatial ambiguous	4521 (2922.71)	4006 (2894.06)	7403 (5022.93)	6808 (8427.04)
Spatial ambiguous	4700 (3034.7)	4414 (5375.04)	10693 (7727.22)	10193 (24998.85)

Table 4:

LMM Results for question reading time (QRT) log(ms) as a function of the presence of ambiguity, Age and stimulus type (ST: non-Spatial/spatial).

Fixed effects	**Question Reading Time (log(RT in ms))**				
	b	**SE**	**df**	**t value**	**Pr(>|t|)**
Intercept	8.59	0.066	190.69	132.012	**<.001**
Age	0.139	0.065	179.66	-2.124	**.035**
ST	0.338	0.057	52.631	5.874	**<.001**
Ambiguity	-0.005	0.019	6665.16	-0.271	.787
Age*ST	-0.025	0.033	6646.22	-0.758	.449
Age*Ambiguity	-0.039	0.034	6657.29	-1.15	.25
ST*Ambiguity	-0.043	0.039	6664.97	-1.084	.750
Age*ST*Ambiguity	0.034	0.048	6657.11	0.712	.476

Figure 3:
LMM estimates for question reading time (QRT) log(ms) as a function of the presence of ambiguity, Age and stimulus type (ST: non-Spatial/spatial). Means shown by red bar



Question accuracy

We further investigated question accuracy. Using Table 5, showing mean accuracy % and standard error, we see again that in accordance with Chapter 2's findings, spatial comprehension accuracy is lower than that of non-spatial accuracy. However, this contradicts the findings of Chapter 4. Furthermore, ambiguity detection is again significantly lower than unambiguous model comprehension. We see no significant differences in the mean comprehension question accuracy. Furthermore, Table 6 shows that ST and ambiguity significantly decrease question accuracy for both old and young and that ambiguous and spatial texts are significantly harder to comprehend than both unambiguous non-spatial and unambiguous spatial texts. Additionally we see no significant differences in question accuracy between older and younger readers, indicating that older and younger participants do not differ in their ability to detect ambiguity and reason from an unambiguous model.

Table 5:

Mean question accuracy by ST, Ambiguity and Age (SD in parenthesis).

Note: General comprehension questions were the same across all conditions.

Younger readers

Type	Mean accuracy (Situation model) %	Mean accuracy (General comprehension)%
Ambiguous non-Spatial	0.391 (0.488)	0.888 (0.315)
Ambiguous Spatial	0.235 (0.424)	0.845 (0.362)
Unambiguous non-Spatial	0.713 (0.452)	0.891 (0.312)
Unambiguous Spatial	0.602 (0.489)	0.852 (0.354)

Older readers

Type	Mean accuracy (Situation model) %	Mean accuracy (General comprehension)%
Ambiguous non-Spatial	0.440 (0.496)	0.892 (0.310)
Ambiguous Spatial	0.245 (0.430)	0.883 (0.322)
Unambiguous non-Spatial	0.721 (0.448)	0.851 (0.355)
Unambiguous Spatial	0.660 (0.474)	0.879 (0.326)

Table 6:

GLMM Results for question accuracy as a function of Ambiguity, ST (non-spatial/spatial) and Age.

Fixed effects	**Question accuracy**				
	b	**SE**	**df**	**z value**	**Pr(>|t|)**
Intercept	1.180	0.200	-	5.89	**<.001**
ST	-0.371	0.211	-	-1.75	.079
Ambiguity	-1.523	0.145	-	-10.45	**<.001**
Age	0.058	0.203	-	0.29	.77
ST*Ambiguity	-0.656	0.205	-	-3.20	**.0013**
ST*Age	-0.242	0.174	-	-1.38	.165
Ambiguity*Age	-0.171	0.177	-	-0.96	.335
ST*Ambiguity*Age	0.407	0.248	-	1.63	.102

Line by Line Reading Times

We investigated how long participant's spent reading each line of the stimulus. We compared the lines that contained a premise (referred to as WasPrem in tables) to those that did not. Here we show that when a line contained a premise they were read and re-read for a significantly shorter amount of time, with the additional caveat that older readers spend significantly longer reading and re-reading ambiguous stimuli. This possibly indicates that older readers have difficulty encoding ambiguous relationships, or that they use a different encoding strategy compared to younger readers. We also find that older readers read both premise and non-premise lines for significantly longer than younger readers, as shown in Table 7. This is consistent with our findings in Chapter 4 as well as what is known about older people's reading behaviour (e.g Rayner et al., 2006). Additionally we found that premise lines are re-read significantly more often by older and younger readers which is shown in Table 8. In turn this finding is consistent with our findings in Chapter 4 and indicates that both older and younger readers are actively aware of which premises require re-reading and return to them to “refresh” their memory (Inhoff, et.al., 2019) prior to moving on to questions.

Table 7:

LMM results for line reading time as a function of premise, and Age.

Fixed effects	**log(Line reading time(ms))**				
	b	**SE**	**df**	**z value**	**Pr(>|t|)**
Intercept	7.98	0.062	172	12700	**<.001**
WasPrem	-0.163	0.014	80300	-13.049	**<.001**
Age	0.281	0.073	15500	4.023	**<.001**
Ambi	0.024	0.012	88300	1.855	.063
WasPrem*Age	0.036	0.025	88300	1.425	.154
WasPrem*Ambi	0.028	0.021	88300	-1.329	.183
Age*Ambi	0.048	0.02	88300	-2.385	**.017**
WasPrem*Age*Ambi	0.034	0.035	88300	0.981	0.3264

Table 8:

GLMM results for re-reading as a function of premise, Age and Ambiguity.

Fixed effects	**Line re-reading likelihood**				
	b	**SE**	**df**	**z value**	**Pr(>|z|)**
Intercept	-0.716	0.119	-	-5.98	**<.001**
WasPrem	0.121	0.028	-	4.31	**<.001**
Age	-0.044	0.197	-	-0.22	.823
Ambi	0.025	0.023	-	1.08	.280
WasPrem*Age	0.005	0.048	-	-0.121	.904
WasPrem*Ambi	0.009	0.039	-	0.217	.829
Age*Ambi	0.065	0.040	-	-1.610	.107
WasPrem*Age*Ambi	0.043	0.067	-	0.651	.515

Model selection

By investigating the answers given by participants when failing to detect ambiguity in ambiguous situations, it was possible to see which of the two viable models was preferred by participants in each item. We separated the unambiguous answers from the unambiguous ones as, when answering questions about unambiguous situations, the reader's choice typically represented the only viable model. This is reflected by the relatively high comprehension accuracy for unambiguous models (Table 5). We calculated how likely participants were to select a model based on premise 1 or premise 2 as when not detecting ambiguity. Table 9 shows the mean selection likelihood across conditions. We ran an LMM analysis in order to determine what influenced the selection of one model over another. As we can see in Table 10, we found no significant differences and only minor model preference for models based on premise 1. Further, older participants were more likely to select specific models (though not significantly so) and both older and younger readers tended to use preferred models in spatial situations. This is visualised in Figure 4, where we can see the minor preference for selecting a model based on premise 1 over premise 2 and that participants did not select the "anchor" often.

Table 9:

Mean likelihood to select premise 1 or premise 2, between premise 1 and 2, if ambiguity is undetected, by ST(spatial/non-spatial) and age group.

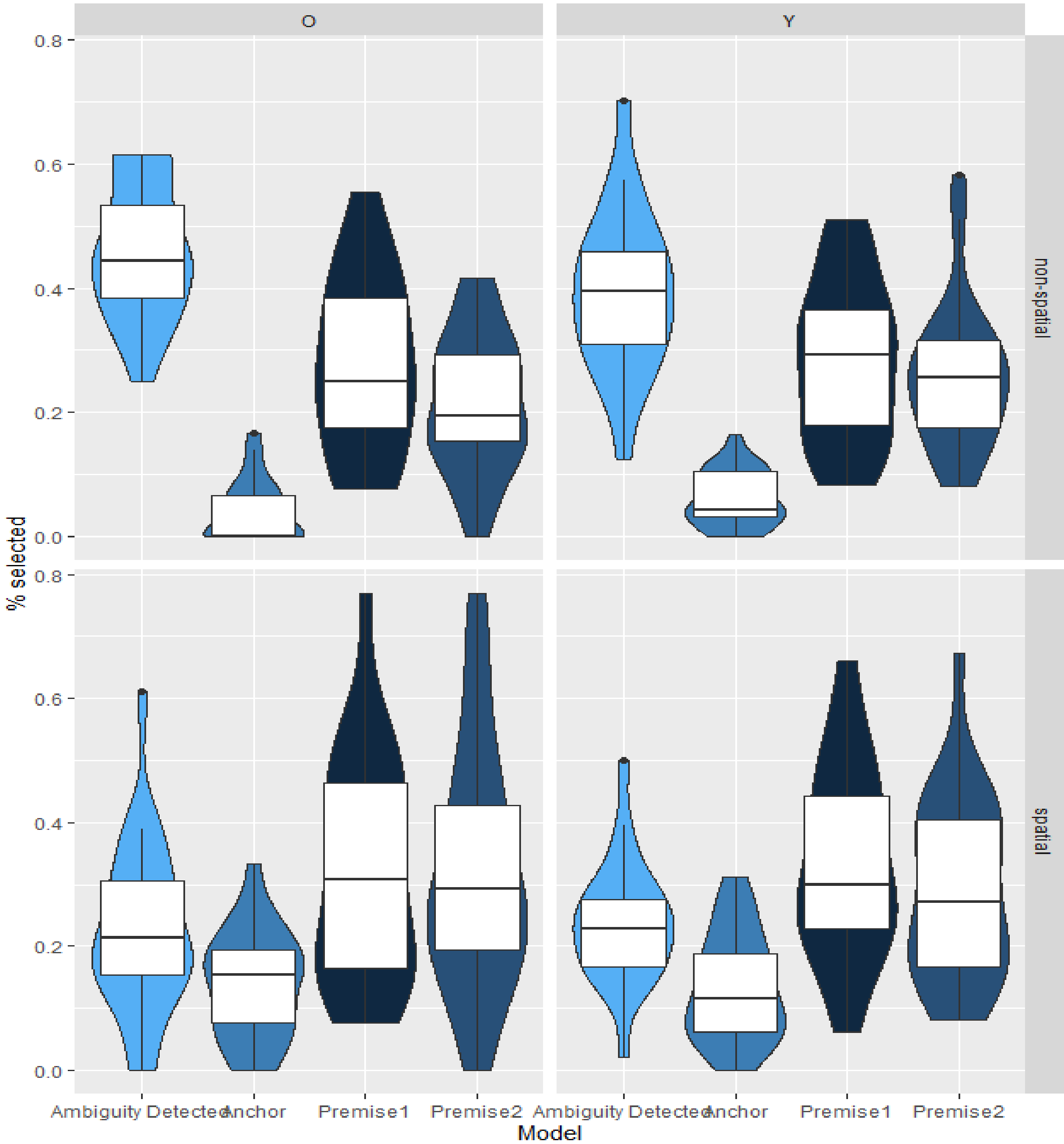
		% premise selection likelihood	
Stimulus Condition	**AgeGroup**	**Selection likelihood of premise 1**	**Selection likelihood of premise 2**
non-spatial	Old	0.565	0.434
non-spatial	Young	0.608	0.391
spatial	Old	0.5	0.5
spatial	Young	0.416	0.5

Table 10:

LMM estimates of likelihood to prefer one model over another as a factor of ST(spatial/non-spatial),Premise (Premise 1 = 0, Premise 2 = 1), and Age Group.

Fixed effects	**Likelihood to prefer one model over another**				
	b	**SE**	**df**	**t value**	**Pr(>|t|)**
Intercept	0.224	0.033	57.40161	6.659	**<.001**
spatial vs non-spatial	0.0601	0.038	42.9519	1.554	0.1276
Premise 2 vs Premise 1	-0.0258	0.034	81.29375	-0.759	0.4498
Young vs old	-0.0481	0.025	32.81926	-1.877	0.0694

Figure 4:Likelihood of model selection by Premise 1 or 2, Ambiguity detection or the Anchor



Discussion of Experiment 1:

Firstly we found no significant differences in which of the two viable models participants preferred in unambiguous situations.While, this goes against prior findings of preferred mental models (e.g Goodwin & Johnson-Laird, 2005; Nejasmic, Bucher & Knauff, 2015) we did find that there are preferences for models based on premise 1, though not significant. However, this finding does support the concept that models are constructed "bit-by-bit" (Johnson-Laird & Byrne, 1991), whereby the object/subject mentioned in premise 2 is "inserted" into the relationship described in premise 1 in a way which does not violate either premise. We also found increased, but not significant, likelihood to select models based on premise 1 in older participants and for spatial situations. It stands to reason that the methodology employed by us in this study is not useful for exploring model preference, as it was not designed for this purpose, rather it is set up for ambiguity detection, spatial and non spatial differences, and changes in situation model processing as we age.

We found that older participants read the simulus significantly slower than younger ones. While this finding is only true for line by line reading, it is not significant for stimulus reading time as a whole. Possibly due to further differences created by re-reading and time spent re-reading. However, this is consistent with what is known about reading changes with age, namely that older readers read more slowly (Rayner 1998, Lowder et al., 2013), and is likely a compensatory mechanism (Rayner et al., 2006) for lower general reading comprehension (Van der Linden et al., 1999; Schurer et al., 2020) . However, we did find both younger and older participants read lines containing premises for significantly less time than non-premises. At first this may seem to be counterintuitive were it not for the fact that both older and younger participants re-read those premise-containing lines of text significantly more often than non-premise lines. It is likely that, as premises are relatively simple to read e.g "apple left of pear", the location of the premises within the text was stored for later "checking" similar to findings in non-premise reading (Inhoff, et.al., 2019). While our findings suggest that older readers are slower overall we do not find a significant increase in re-reading likelihood for older readers which goes against the well established risky reading hypothesis (Rayner et at., 2006) as well as our previous findings. It is likely that our findings are different from work on risky reading due to the limitations of this study, as we cannot assess the eye-movement related reading behaviour of readers. Alternatively, it may be that both younger and older readers are reading the texts more carefully than they usually would and are therefore attempting to better encode important information; we see evidence of this in the higher re-reading rate for premise lines.

We investigated QRT and found that the ST significantly increased QRT, namely participants took longer to answer questions about spatial situations. However, we found that the overall differences between the spatial and non-spatial comprehension was not significantly different, replicating our finding from Chapter 4. Further, in replication of our findings of Chapters 2 & 4, spatial situations are harder to reason out than non-spatial ones as shown by the significantly longer question reading/reasoning time. We also found that older readers had a significantly longer question reading time than younger readers, However, we find no significant differences in model comprehension accuracy between young and old, consistent with what is known about mental model reasoning in ageing (e.g.,Gilbert et al., 2004;Radvansky et al., 1990). A possible explanation for this is that older readers spend more time reading the question in order to compensate for age-related cognitive declines, possibly trying to inhibit irrelevant information (Hasher & Zacks, 1988; Hamm & Hasher, 1992) from memory (Gerard et al., 1991) when answering the question, or trouble with integrating the premises (Copeland & Radvansky, 2007). In any case this strategy pays off as we failed to find significant differences in question accuracy between older and younger readers. In fact, older readers had slightly higher average question accuracies, though this is not a significant difference. Our findings here are consistent with prior work (e.g., Radvansky et al., 1990), in that older readers only seem to struggle with memory retrieval but not the creation and reasoning from a mental model (Gilbert et al., 2004). Furthermore, it appears that older and younger readers are hindered by spatial ST and ambiguity and perform significantly worse in these conditions. Though neither old or young performs significantly differently from the other.

Lastly, in our initial exploration of situational ambiguity we suggested that ambiguity is not realised during the reading of the stimulus but rather during the reading time of the question as our initial findings showed an increase in QRT but not SRT for ambiguous situational models. It appears that our initial suggestion was incorrect or incomplete as our findings here suggest that QRT is not significantly different for ambiguous situations (see table 4), while question accuracy is significantly lower for ambiguous situational models. Despite lower question accuracy being indicative of comprehension difficulty, we cannot attribute this to ambiguity being realised only during question reading. Furthermore, table 8 shows that re-reading increases significantly for individual lines containing a premise. It remains likely that situational models are processed and stored “bit by bit” as has been found previously (Johnson-Laird & Byrne, 1991), though it is not possible to ascertain when ambiguity is detected. Potential future work may benefit from re-evaluation tasks (e.g Rauh et al., 2005) in order to investigate the process underlying ambiguity detection.

For ambiguous models it is possible for a question to look as though it probes a single premise i.e “A in relation to B” or “Is A next to B” (external). However, when the other premise is taken into account, the question actually probes the whole model i.e “is this model ambiguous?” or “Is C between A and B?” (internal). Therefore these questions may encourage readers to not consider the second premise. Naturally, the use of only a single element of the model would lead readers to not detect ambiguity, which would be consistent with our results. For both older and younger readers, we found higher re-reading rates and longer reading time for lines containing the premises but found no significant differences for ambiguous stimuli. It may therefore be that ambiguity in the model offers no additional difficulty in processing compared to an unambiguous model, similar to what has been found in processing of syntactically unambiguous and syntactically ambiguous sentences (van Gompel et al., 2005).

It is difficult to determine when readers realise a described situation is ambiguous. Regardless of when this occurs, the detection of ambiguity is only possible when two premises are integrated into a single model. Such an integration may be too difficult to process or memorise at which point a “good enough” or lazy model is opted for instead. It is possible that a “good enough” interpretation of one of these premises could lead a reader to assume the model is unambiguous. However, at present, this remains a speculation.

Experiment 2

In this experiment we investigate if probing the “internal” part of the mental model leads to better ambiguity detection and model comprehension. If readers create a “lazy” model from any read situation or if a preferred model is used to make inferences about an ambiguous situation we assume that this hindered both model comprehension and ambiguity detection. When a model is constructed in such a way, accurate answers should rely equally on both premises. The Experimental design, Stimuli, Procedure and Analysis are exactly the same as Experiment 1 with the only difference being the wording of the questions.

2: Participants

We collected data from 47 young (Mean age= 31.91 years, SD=8.96 years, 34 Female) and 52 old (Mean age=65.83 years, SD= 4.91years, 32 Female) above the age of 60 native English speakers , recruited online from Prolific.co. All participants were naïve as to the ambiguity manipulation, though were instructed to answer “There is not enough information” if there was not enough information (i.e Ambiguous) and were informed that they would be reading about non-Spatial and spatial situations. The study was approved by the Bournemouth University Research Ethics Committee (ID 27563).

Materials and Designs

The stimuli consisted of 48 paragraphs: 24 describing a spatial situation and 24 describing a non-Spatial one. Each of these paragraphs could be seen in one of two versions: ambiguous or unambiguous. The stimuli were identical to those used in experiment 1. However, as mentioned we changed what the question probed and as a result the questions have different numbers of words. Table 9 shows the descriptive metrics of the questions. In contrast to Experiment 1 the questions in this experiment probed the “internal” part of the model/s by asking “ Which of the objects/subjects is between the two others?”.

Table 9: Descriptive information of questions probing the “internal” part of the model used in Experiment 2

Condition	**Ambiguity**	**Word count**	**Word Length in letters**	**Flesch-Kincaid**
Spatial	Yes	8.54 (1.64)	4.51 (2.06)	4.14 (1.28)
Spatial	No	8.54 (1.64)	4.51 (2.06)	4.14 (1.28)
non-Spatial	Yes	9.65 (1.13)	4.63 (2.27)	4.38 (1.64)
non-Spatial	No	9.65 (1.13)	4.63 (2.27)	4.38 (1.64)

2: Results

We removed 7 participants due to low accuracy (below 70%) in general comprehension questions. Following the same elimination protocols as in Experiment 1, 93.7% of the data was left over for Analysis.

2: Reading time of Stimulus text

We again began with the investigation of SRT in their single paragraph format in order to replicate the finding from our previous study that there were no significant differences between the reading times of non-Spatial, spatial, ambiguous and unambiguous paragraphs.

Table 10

Mean stimulus reading times in ms (SD in parenthesis).

	Stimulus RT (ms)	
Stimulus Condition	**Young**	**Old**
non-Spatial unambiguous	47270 (42154.83)	46785 (27879.75)
Spatial unambiguous	46629 (26926.86)	50142 (33290.14)
non-Spatial ambiguous	46994 (38291.50)	45278 (24157.06)
Spatial Ambiguous	45357(28652.18)	47263 (32544.17)

Table 10, shows means only, however we investigated stimulus reading time and found only null effects. These are available to view as part of the analysis code on the OSF which can be accessed via the link on the cover page.

2: Question reading time

We investigated the QRT of older readers and compared these to the QRT of younger readers. The means and standard deviations can be seen in Table 11. As we found in Chapter 4 and in Experiment 1 We find that QRT is significantly higher for older readers. We also found that both older and younger readers spend significantly more time reading spatial questions than non-spatial questions, which was found in Chapter 4 and in Experiment 1. However, in contrast to Experiment 1, in this experiment we found that older readers read questions about spatial situations for a significantly shorter amount of time than younger readers. This could imply that probing internal questions benefits spatial reasoning in older readers but not in younger readers.

Table 11: *Mean Question reading times in ms and mean comprehension question reading time(SD in parenthesis).*

	Question RT (ms)			
	Comprehension Question		Situational Question	
Question Condition	**Old**	**Young**	**Old**	**Young**
Non-spatial unambiguous	4645 (3343.89)	4144 (2489.05)	7527 (1144.99)	4668 (4761.79)
Spatial unambiguous	5260 (9761.68)	4512 (272354)	8164 (1357.39)	5743 (4778.70)
Non-Spatial ambiguous	4831 (4172.75)	4227 (2795.48)	8338 (1647.99)	5266 (13132.82)
Spatial ambiguous	4987 (3493.72)	5519 (25491.28)	8223 (1647.99)	6198 (6150.91)

Table 11:

LMM Results for question reading time (QRT) log(ms) as a function of Age, the presence of ambiguity, and stimulus type (ST; non-Spatial/spatial).

Fixed effects	**Question Reading time (log(RT in ms))**				
	b	**SE**	**df**	**t value**	**Pr(>|t|)**
Intercept	8.19	0.062	183	130.22	**<.001**
Age	0.37	0.080	154	4.69	**<.001**
ST	0.228	0.047	114	4.78	**<.001**
Ambiguity	-0.020	0.033	5420	-0.59	.55
Age*ST	-0.109	0.046	5420	-2.35	**.0187**
Age*Ambiguity	0.004	0.046	5430	0.09	.925
ST*Ambiguity	0.0403	0.047	5420	0.85	.394
Age*ST*Ambiguity	0.0018	0.065	5420	0.02	.977

2: Question Accuracy

We again investigated question accuracy. Table 12 shows mean accuracy % and standard deviation. The results of a GLMM are shown in Table 13. Here we found, in line with all previous Chapters and Experiment 1, a significantly lower accuracy, for both young and old, when answering questions about ambiguous situations relative to non-ambiguous questions. As with Experiment 1 and Chapter 2 we found a significantly lower comprehension accuracy for spatial questions compared to non spatial ones. In contrast to our predictions, but in line with the findings of Chapter 4 and experiment 1, while we found no significant age related differences in overall question accuracy. However, older readers have significantly higher accuracy than younger readers when answering spatial questions. This may reflect a similar effect to the decreased QRT in older readers for spatial questions. This again indicates that older readers appear to have benefitted from probing the internal part of the model more so than younger readers. Additionally we found that older readers are significantly better at detecting ambiguity than younger readers. This finding does not go directly against our previous predictions, as we speculated that probing the internal part of the model would aid in ambiguity detection. However, we assumed that probing the internal part of the model would benefit older and younger readers equally, which does not appear to be the case as older readers outperformed younger readers in ambiguity detection.

Table12:

Mean question accuracy of older readers by ST and the presence of ambiguity and Age (SD in parenthesis).

Old

Type	**Mean accuracy (Situational) %**	**Mean accuracy (comprehension)%**
Ambiguous non-Spatial	0.616 (0.486)	0.877 (0.328)
Ambiguous Spatial	0.501 (0.500)	0.886 (0.318)
Unambiguous non-Spatial	0.702 (0.457)	0.890 (0.312)
Unambiguous Spatial	0.685 (0.464)	0.874 (0.332)

Young

Type	**Mean accuracy (Situational) %**	**Mean accuracy (comprehension)%**
Ambiguous non-Spatial	0.579 (0.494)	0.879 (0.325)
Ambiguous Spatial	0.417 (0.493)	0.844 (0.363)
Unambiguous non-Spatial	0.739 (0.439)	0.863 (0.344)
Unambiguous Spatial	0.642 (0.479)	0.842 (0.364)

Table 13:

GLMM Results for question accuracy as a function of the presence of ambiguity, ST (non-Spatial/spatial) and Age.

Fixed effects	**Question accuracy**				
	b	**SE**	**df**	**z value**	**Pr(>|z|)**
Intercept	1.38	0.236	-	5.85	**<.001**
ST	-0.568	0.233	-	-2.43	**.0149**
Ambiguity	-0.958	0.139	-	-6.92	**<.001**
Age	-0.253	0.268	-	-9.45	.344
ST*Ambiguity	-0.322	0.191	-	-1.69	.09
ST*Age	0.516	0.191	-	2.69	**.007**
Ambiguity*Age	0.459	0.190	-	2.41	**.0159**
ST*Ambiguity*Age	-0.276	0.264	-	-1.05	.295

2: Line by Line Reading Time

We investigated how long readers spent reading each line of the stimulus. We compared the lines that contained a premise to those that did not. We again see that premise lines were initially read for a significantly shorter amount of time than non-premise lines, as seen in Table 14. We also see that older readers read for significantly longer, however we failed to find significant differences in how long older readers read premise lines. We also investigated the re-reading rate of individual lines. In table 15 we see that older readers reread lines significantly more often than younger readers, this may reflect the prior findings in terms of question accuracy and QRT whereby internally focused questions aid older readers in model processing. However, we also found that both older and younger readers read premise lines more often than non-premise lines.

Table 14:

LMM results for line reading time as a function of premise, and Age.

Fixed effects	**log(Line reading time(ms))**				
	b	**SE**	**df**	**t value**	**Pr(>|t|)**
Intercept	7.64	0.061	113	123	**<.001**
WasPrem	-0.155	0.021	5670	-7.36	**<.001**
Age	0.332	0.080	98.7	4.13	**<.001**
Ambi	-0.005	0.017	5670	-0.31	0.756
WasPrem*Age	0.002	0.030	5670	0.066	0.948
WasPrem*Ambi	0.016	0.029	5670	0.558	0.577
Age*Ambi	0.011	0.024	5670	0.449	0.653
WasPrem*Age*Ambi	-0.019	0.042	5670	-0.464	0.643

Table 15:

GLMM results for re-reading as a function of premise, and Age.

Fixed effects	**Line re-reading likelihood**				
	b	**SE**	**df**	**z value**	**Pr(>|z|)**
Intercept	-0.854	0.146	-	-5.86	**<.001**
WasPrem	0.054	0.041	-	1.34	.180
Age	0.527	0.213	-	2.48	**.0132**
Amb	0.073	0.034	-	2.11	**.0348**
WasPrem*Age	0.055	0.056	-	0.97	.332
WasPrem*Amb	0.11	0.057	-	1.75	.081
Age*Amb	-0.01	0.047	-	-0.17	.869
WasPrem*Age*Amb	-0.071	0.078	-	-0.89	.369

2: Model selection

We again investigated the preferred models of participants when failing to detect ambiguity. As in Experiment 1 we omitted the unambiguous answers as these had relatively high comprehension accuracy, shown in Table 13. In contrast to Experiment 1, where we found no significantly higher likelihood to select one or the other model, we found that the likelihood to select one particular model over another is significantly higher in spatial situations. However, in this case the likelihood is reversed whereby a model based on Premise 2 was selected more often than one based on Premise 1, but as in Experiment 1 this proclivity is not significant. In another reversal of Experiment 1 we found that younger readers were more likely to select one model over another, but again this was not significant. A potential interpretation of this reversal is, as we have suggested previously, that asking internally probing questions is a significant boon to older readers, which may cause them to detect ambiguity more often and not create a preferred or lazy model. However, if this idea is to be applied to model selection, a more detailed analysis (which follows shortly). Lastly, Figure 5 visualises the likelihood to select a particular model over another. Figure 5 shows that selecting the correct answer “there is not enough information” is much higher than selecting any of the other types of model, and that selecting any of the models are below chance level which may indicate that the previously mentioned effects in model selection are not as informative as they first appear to be.

Table 16:

Mean likelihood to select premise 1 or premise 2, between premise 1 and 2, if ambiguity is undetected, by ST(spatial/non-spatial) and age group.

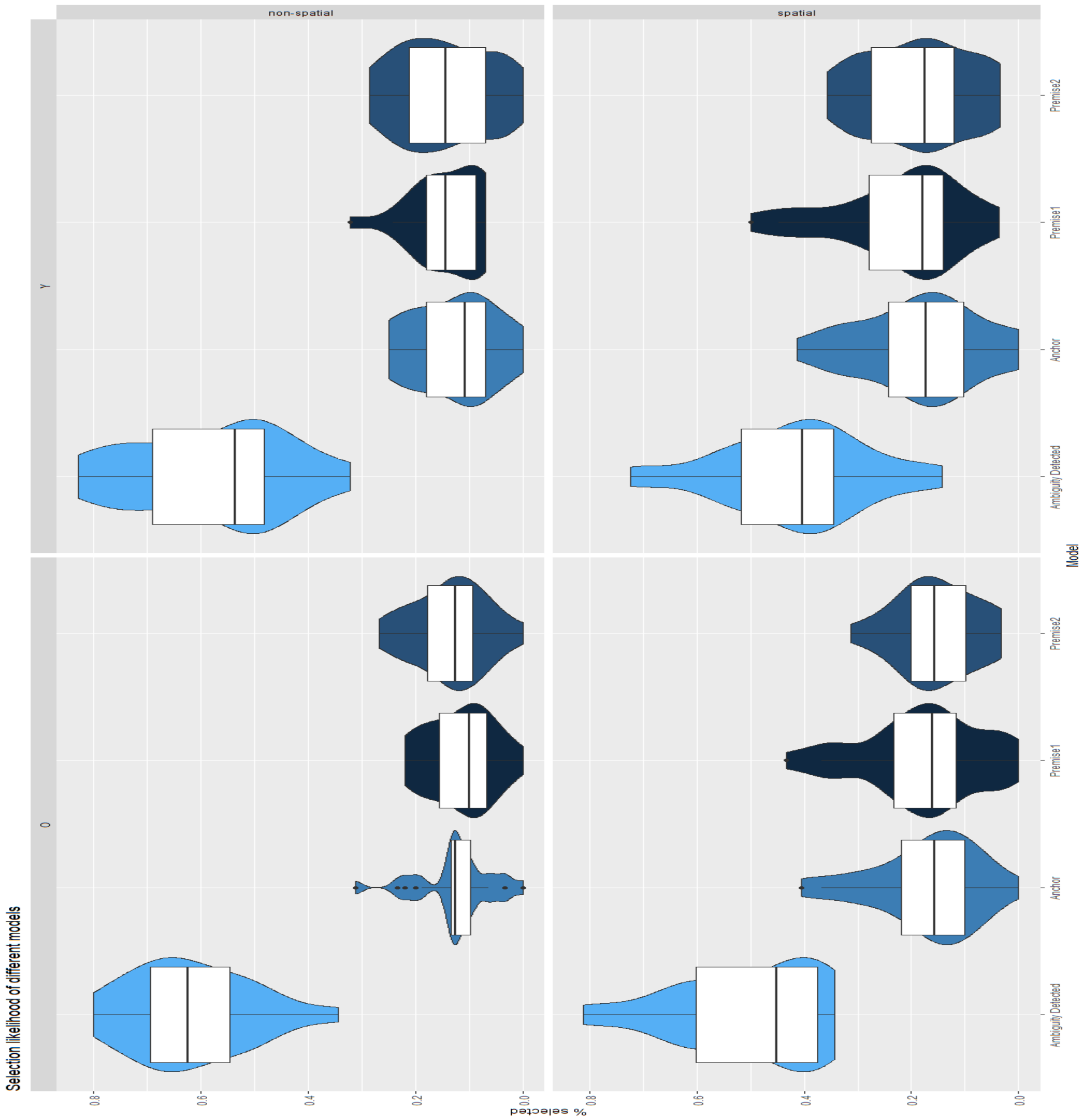
		% premise selection likelihood	
Stimulus Condition	**AgeGroup**	**Selection likelihood of premise 1**	**Selection likelihood of premise 2**
non-spatial	Old	0.565	0.434
non-spatial	Young	0.478	0.521
spatial	Old	0.583	0.416
spatial	Young	0.708	0.291

Table 17:

LMM estimates of likelihood to prefer one model over another as a factor of ST(spatial/non-spatial),Premise (Premise 1 = 0, Premise 2 = 1), and Age Group.

Fixed effects	**Likelihood to prefer one model over another**				
	b	**SE**	**df**	**t value**	**Pr(>|t|)**
Intercept	0.0647	0.016	78.507	4.009	**<.001**
spatial vs non-spatial	0.0386	0.016	44.947	2.29	**.0267**
Premise 2 vs Premise 1	0.0197	0.016	88.400	1.216	.227
Young vs old	0.0203	0.014	45.53	1.382	.173

Figure 5: Selection likelihood violin plot showing the difference in participants selecting a model based on Premise 1, Premise 2, the anchor or answering the question correctly.



Discussion Experiment 2:

In Experiment 2, we investigated how discouraging the creation of “lazy” models and revision of preferred models might influence the detection of ambiguity and model comprehension, the key results are as follows:

In line with all previous Chapters and Experiment 1 we again found that ambiguity detection for both young and old readers was significantly lower than unambiguous model comprehension. In contrast to our predictions, but in line with the findings of Chapter 4 and Experiment 1, we found no significant age related differences in overall question accuracy. It is therefore highly likely that ageing does not influence the ability of an older person to create, and reason with a mental model, which has been found before (Brébion, 2003), despite difficulty in memory retrieval (Gerard et al., 1991).

However, in contrast to Experiment 1, where we found no significant differences between older and younger readers in ambiguity detection, we found that in Experiment 2 older readers appeared to be significantly better at detecting ambiguity. Older participants outperforming younger ones, in terms of mental imagery has been found before (e.g. Dror et al., 2005), where older adults performed significantly at rotating complex mental icons. The findings of Dror et al. indicated that older adults tended to adopt a more holistic approach in order to reduce cognitive load. It stands to reason that a more holistic strategy is also used by older readers when making inferences about mental models and that probing the internal part of the model aids people using this strategy in some way. Potentially, the proclivity of older participants to break down information into smaller “chunks” (Swets et al., 2007) and the “bit-by-bit” construction of models (Johnson-Laird & Byrne, 1991) make using a “holistic” approach to model construction (Dror et al., 2005), using both premises to construct the model, an especially good strategy to detect ambiguity. Further, subsequent model revisions (Rauh et al., 2005) may take place at the “holistic” level rather than only in part of the model. It stands to reason that this process may be somewhat lengthy, which would explain the significantly higher question reading/reasoning time for older participants compared to younger ones. However, as older readers are generally slower readers/reasoners (e.g. Myerson et al., 1990) it is not possible to determine if this is the case.

Older readers had significantly higher accuracy than younger readers when answering spatial questions, which again indicates that older readers appear to benefit from internal model probing for more than just ambiguity detection. This finding does not go directly against our previous predictions, as we speculated that probing the internal part of the model would aid in ambiguity detection. However, we assumed that probing the internal part of the model would benefit older and younger readers equally, which does not appear to be the case. It is known that ageing does not affect a person's ability to create a mental model (Radvansky et al., 1990), furthermore it is possible that certain strategies, used by older readers to reduce cognitive loads e.g (Dror et al ., 2005) may actually be beneficial in solving/creating the spatial situational models we create through texts. Older participants' higher spatial reasoning compared to younger participants may be due to a more holistic approach. It is possible to speculate that by probing the internal part of the model, older readers are better able to process spatial configurations using both premises. While this study does not investigate the particular reasoning strategy, it is possible that probing internal parts of the model may require a different strategy compared to probing the external part of the model. Furthermore, it is known that strategies for processing mental models do differ from person to person (Schaeken et al 2000) and older people are known to use compensatory strategies/mechanisms (described in Piefke et al., 2012) e.g., "risky reading (Rayner et al., 2006). Such strategies used by older participants may take more time (Gilbert et al., 2004) but yield similar text comprehension (Radvansky et al., 1990).

We also see several significant differences in older and younger participants reading behaviours during stimulus presentations. Both older and younger readers reread premise lines more often than non-informative ones, likely to better enshrine the information into memory (Margolin & Snyder, 2017; Rawson, et.al ,.2000; Stine-Morrow et.al ,. 2004). When compared to one another older readers are significantly more likely to re-read informative lines than younger readers. However, it is difficult to determine if this is due to "risky reading" (Rayner et al., 2006) which makes re-reading more prevalent, lower working memory capacity (Bopp & Verhaeghen, 2007; Borella et al., 2008; De Beni et al., 2004; Fiore et al., 2012 ; Park et al., 2002) or because the text is harder for older readers to interpret (Inhoff, et.al., 2017). However, when a line contained a premise they were read for a significantly shorter amount of time. For both older and younger readers this behaviour could indicate a type of search for relevant information, once the premise is found it isn't read for particularly long as the information is more easily extracted once located but return to it more often in order to memorise and clarify the information. Though this is not possible to say for certainty without the use of eye tracking.

Lastly, we found few significant differences in what kind of model readers prefer when ambiguity was not detected. However, we found that one model was significantly preferred over another in spatial situations and this may be influenced by an semantic interference on spatial models (Knauff & Johnson-Laird 2002;Knauff & May, 2006) which we have previously suggested was the cause for lower comprehension accuracy of spatial situations and lower rates of ambiguity detection. On the other hand, it is important to note that Experiment 2, for one reason or another, has significantly higher rates of ambiguity detection and a slightly smaller sample size compared to Experiment 1. Therefore the significantly higher model preference likelihood in spatial situations may not be wholly accurate.

Internal/External Comparison

In order to investigate the idea that externally/internally probed question types change the way in which people interpret situational models, we compared QRT and Question accuracy between the two experiments. We also investigated model preferences between the two experiments. Lastly we investigated these through the lens of ageing as a main factor in the models, as probing internal models may be more beneficial for older readers.. It is important to note that there is likely to be some variation between the two experiments as they were based on different samples, as well as sample sizes. Reading/reasoning time and question accuracy may shed light onto whether or not ambiguity detection is task dependent, or whether specific lines of questioning improve participant's comprehension of the model. Furthermore, by investigating this through the lens of ageing it is possible to investigate if specific lines of questioning aid older readers, either through a reduced working memory load or model clarity. As the stimuli did not change between the two experiments we did not investigate any line by line differences between internal and externally probing questions. However, it should be noted that such an investigation may bear fruit as reading behaviours may be influenced by the task which they precede.

3: Question reading time

We investigated how QRT differed between the two studies. We found that internal questions took significantly less time to answer. However, the difference in reading/reasoning time between the internal and external groups is likely due to the significantly higher question word count and lower readability in Experiment 1. Because of this confound we do not present any of the question reading time analysis, however the analysis and differences in word count and readability between questions are included in the supplementary materials for this Chapter.

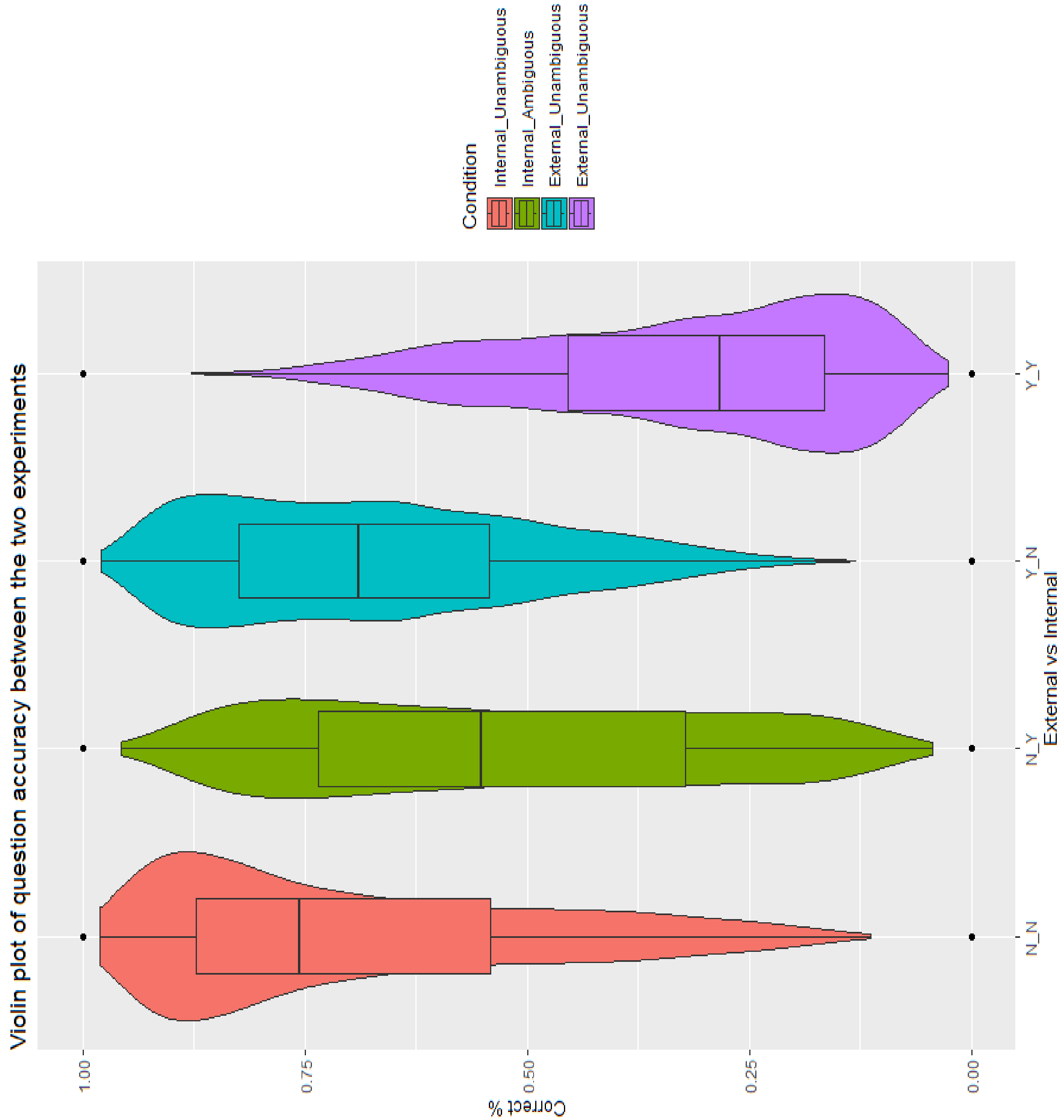
3: Question Accuracy

We investigated the differences in question accuracy between the external and internal groups. In table 17 we see a GLMM comparison between the group that had internal and external question types (external 0 , as in experiment one & internal 1, as in experiment two). Across both studies the accuracy effect of Subject Type is pervasive where Spatial models have significantly lower comprehension accuracies. However, the low comprehension accuracy for spatial situations and low ambiguity detection is nothing new as we have found them in Chapters 2 and 4 as well as Experiments 1 and 2 in this chapter. However, while hard to detect overall, ambiguity is significantly easier to detect when questions probe the internal aspect of the model. In regard to older participants benefiting from internally probing questions we found no significant differences in comprehension accuracy between the two experiments. There are no major interactions with the internal/external question manipulation. The significant difference in ambiguity detection between the two experiments is visualised in Figure 6.

Table 17:
GLMM results for Question accuracy as a function of Persp (Internal/External) interacting with Age group, ST, and the presence of ambiguity, between the two groups.

Fixed effects	**Question Accuracy**				
	b	**SE**	**df**	**z value**	**Pr(>|t|)**
Intercept	1.08	0.123	-	8.79	**<.001**
Persp	0.207	0.206	-	1.01	.313
Age	0.002	0.213	-	0.010	.992
Ambi	-1.61	0.099	-	-16.2	**<.001**
ST	-0.584	0.96	-	-6.06	**<.001**
Persp*Age	-0.232	0.311	-	-0.748	.454
Persp*Ambi	0.726	0.166	-	4.38	**<.001**
Age*Ambi	0.245	0.169	-	1.45	.147
Persp*ST	0.025	0.164	-	0.153	.879
Age*ST	0.258	0.167	-	1.54	.123
Ambi*ST	-0.269	0.139	-	-1.94	.053
Persp*Age*Ambi	0.164	0.249	-	0.659	.510
Persp*Age*ST	0.206	0.248	-	0.828	.408
Persp*Ambi*ST	-0.009	0.229	-	-0.0410	.967
Age*Ambi*ST	-0.406	0.238	-	-1.71	.088
Persp*Age*Ambi*ST	0.183	0.346	-	0.528	.598

Figure 6:
Differences in questions accuracy between Internal (Experiment 2) and External (Experiment 1) Questions.



3: Model selection

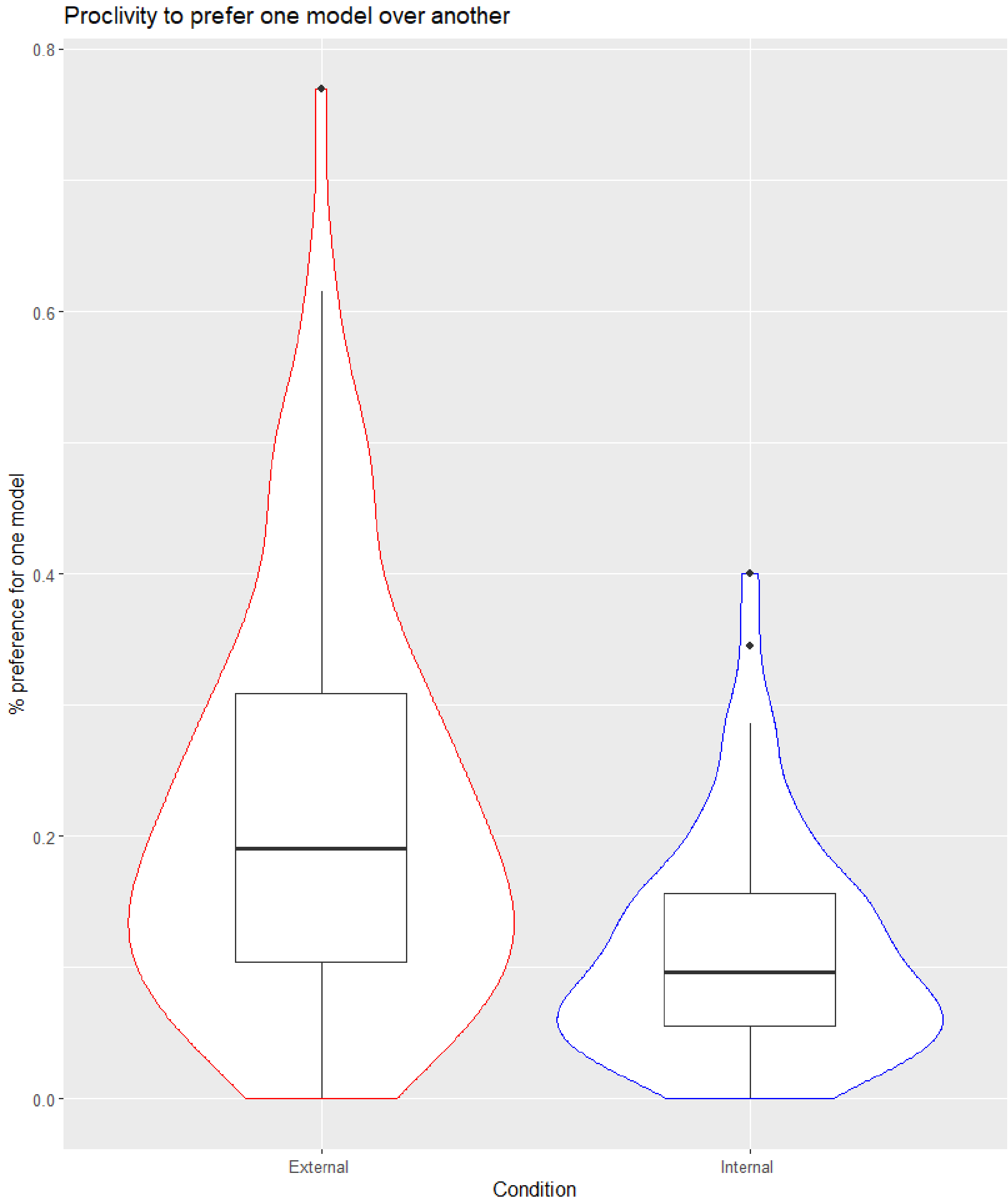
We compared the likelihood for participants to prefer one model over another. As before we only investigated the model preference likelihood for ambiguous stimuli. We can see from Table 18 that there are no significant differences as to which model was preferred over another. However, we see that preferring one model over another model is significantly lower when probing the internal part of the model, indicating that probing the internal part of the model discourages the use of a lazy model. On the other hand this may also indicate that preferred models are revised upon questioning and ambiguity is detected significantly more often (as shown in Table 17).

Table 18: Comparative likelihood to select a model based on one premise over another as a factor of which premise had a higher selection likelihood (Premise 1vs Premise 2), Internal/External probing questions and AgeGroup

Fixed effects	**Selection likelihood lean**				
	b	**SE**	**df**	**t value**	**Pr(>|t|)**
Intercept	0.2562	0.026	110.450	9.715	**<.001**
Premise 2 vs Premise 1	-0.026	0.036	136.983	-0.723	.471
Internal vs External	-0.161	0.037	109.994	-4.277	**<.001**
Young vs Old	-0.041	0.033	103.066	-1.252	.213
Premise 2*Internal	0.0303	0.056	136.535	0.536	.593
Premise 2*Young	0.0017	0.048	115.631	0.036	.971
Internal*Young	0.0476	0.050	110.755	0.938	0.35
Premise 2*Internal*Young	0.0243	0.077	130.184	0.313	.755

Figure 7:

Violin plot showing preference to select one model over another, Internal vs. External question types.



3: Discussion of the internal/external comparison

We found a significant increase in the reading time of older readers across both Experiments for question reading time and line by line reading time. Note that here we are referring to line by line reading time which we use to reflect the stimulus text reading time overall. This slower reading time is something that appears to be common in ageing and reading research and may be an age-related decrease in inhibiting irrelevant information (Hamm & Hasher, 1992 ; Hasher & Zacks, 1988) or declines in working memory (Dror et al., 2005). To make a comparison between the two experiments in this case would create a confound as the two experiments are based on two different participant groups. However, a comparison of this reading time difference can be found in the appendix for this chapter and shows no significant differences between the two experiments in terms of line by line and is only an age related difference.

We find that across both studies that older readers had significantly higher QRT, again we did not directly investigate the difference between the two experiments due to the two different groups of participants. Age-related slowing in reading speed is known (Kemper et al., 1993; Liu et al., 2017) but it is not possible to discern reading speed from reasoning speed in this experiment. It is possible that this age-related difference in QRT is due to difficulty in memory retrieval (Gerard et al., 1991) but it is also possible that a different reasoning strategy is used (e.g, Dror et al., 2005; Schaeken et al 2000) which incurs higher time costs (e.g., Copeland & Radvansky, 2007; Gilbert et al., 2004). Further, evidence to support a different strategy being used can be found in the higher re-reading rates of older participants as well as the differences in model comprehension in Experiment 2.

We found that in both experiments, older readers in general have a tendency for older readers to re-read more often than younger readers. It is generally accepted that older readers reread more than younger readers (Li et al., 2019; Rayner et., 2006). As stated it is likely that this comes along with age related increases in re-reading potentially an adaptive strategy to negate declines in memory (e.g Inhoff, et.al., 2019; Wong & Moss, 2021). We also found significantly higher QRT for spatial texts in both experiments, which is consistent with the findings in our previous empirical Chapters. In relation to ambiguity we found that, while older readers had a higher likelihood of re-reading overall, both old and young participants had a significantly higher likelihood to re-read ambiguous compared to unambiguous stimuli. This indicates that both older and younger readers may be aware that the situation described is ambiguous during reading. Although based on different samples this may indicate that the question type manipulation in Experiment 2 is beneficial to the detection of ambiguity, as the higher re-reading likelihood in Experiment 2 is not found in Experiment 1. Re-reading behaviours appear to be beneficial for both older and younger readers, in ambiguous and unambiguous texts. This could indicate that ambiguity detection is deterministic in nature, and highly dependent on the task following it. We suggest the manipulation in Experiment 2 aids both older and younger readers to detect ambiguity and that if ambiguity is easier to detect re-reading is more likely as participants ensure that the model/s they construct are accurate. We can see evidence for this in the difference in question accuracy between the two experiments and re-reading differences present in Experiment 2 but not in Experiment 1. Furthermore, by encouraging both older and younger readers to deal with the model more "holistically" through the use of an internal model probe, we have shown that though people may be "blind" to alternate models when performing a model construction task (e.g., Rauh et al., 2005) they are not incapable of maintaining multiple functional models in working memory.

Additionally, in Experiment 2 we found that older readers were significantly faster than younger readers when reading/processing spatial texts. This may make a case that older reader's risky reading strategy in combination with a discouraged use of a "lazy model" and/or making inferences from a single preferred model, allows them to integrate spatial models much faster than younger readers. This is further supported by the significant increase in spatial comprehension for older readers compared to younger ones. A higher spatial model comprehension in older participants compared to younger participants has been found before (e.g, Dror et al., 2005) and a similar model processing ability in young and old is not unusual either (Radvansky et al., 1990). Furthermore, we do not see this effect in Experiment 1 and so it could be that this is a result of the question manipulation being a bigger boon to older readers than younger ones. We speculate that probing the “middle” of the model may in fact allow participants to better integrate the two premises together in a more holistic manner and while this is beneficial for both old and young (in terms of model comprehension accuracy and ambiguity detection compared to probing the external part of the model). However, the greater speed at which older participants reason/read spatial questions compared to younger readers may be down to the fact that they already prefer to deal with spatial models holistically (Dror et al., 2005) due to a decline in working memory as they have aged (Brébion, 2003). It has to be noted that in general older and younger readers do not generally differ in their ability to make inferences about spatial or non-spatial models which is known from Chapter 4 and Experiment 1. It is also known that there are no age related differences in the ability to make inferences about models (McKoon & Ratclif, 2013).

Both older and younger participants struggle with comprehension of spatial models when compared to non-spatial ones. While we have previously suggested that the difficulty in making inferences about spatial models, and detecting ambiguity, may come from an increased semantic interference (e.g., Knauff & Johnson-Laird 2002;Knauff & May, 2006). In Experiment 2 we found some evidence to suggest that semantics leads participants to prefer one model over another but only in spatial situations. Furthermore, while not significant, we found an overall increase in preferring one model over another for spatial situations in Experiment 1. Both of which indicate that readers of spatial models may be more inclined to prefer one interpretation over another when reading ambiguous situations. The stimuli in this study are not well suited for making assertions about model preference but similar findings to prior work investigating mental models (Franz & Garnham, 1990; Baddeley, 1998;

Johnson-Laird & Byrne, 1991; Rauh et al., 2005; Goodwin & Johnson-Laird, 2005; Knauff & May, 2006) would suggest that one model is preferred.

An alternative explanation would be that problems with scene visualisation (Ford, 1995; Bacon et al., 2008) can also interfere with how a model is constructed by the participant. For example a non-spatial model can be solved along any axis that the reader preferes i.e "the fastest runner" can be placed anywhere e.g "furthest to the right" in the mind's eye". Spatial situational models are more likely to be bound to the axis in which they are presented, if this model is ordered in another way, the additional cognitive step of rotation would be required. While it is not known if participants switched the axis in this task a reader may prefer this strategy (Schaeken et al., 2000). This may increase the cognitive resources required for spatial models but wouldn't be needed in non-spatial models, as the latter was likely already created along the reader's preferred axis. It is possible that the increase in QRT for spatial questions in both experiments is the result of this higher cognitive load. However, we found no significant differences in spatial model comprehension accuracy between experiments 1 & 2, which indicates that removing the need for a specified axis (the middle of a model is always between the other two no matter the angle from which it is viewed), does not improve model comprehension. Therefore the difficulty in spatial reasoning is unlikely to be due to preferred axis' or perspectives as was one of our suggestions in Chapter 2.

Lastly, we find few significant differences between old and young participants and the two experiments in terms of which model was preferred. It is known that model construction strategies differ from person to person (Bacon et al., 2003) and so it is likely that not everyone would prefer to create their preferred model in a particular way. However, we did find that there is a non-significant proclivity for older participants and all participants making inferences about spatial situations to prefer a model constructed based on Premise 1. A finding which is consistent with a model being constructed "bit-by-bit" (Johnson-Laird & Byrne, 1991). It is likely that due to the two Experiments being geared toward ambiguity detection and not model preference detection that the effects are somewhat weakened in comparison to prior work on mental models (e.g Franz & Garnham, 1990; Baddeley, 1998; Johnson-Laird & Byrne, 1991; Rauh et al., 2005; Goodwin & Johnson-Laird, 2005; Knauff & May, 2006), which find that reasoners generally prefer a single model.

Conclusion

In conclusion we found that age related cognitive declines do not actively hinder the detection of ambiguity nor the solving of situational models overall, though consistent with prior research older participants took longer to read and reason (Copeland & Radvansky, 2007; Gilbert et al., 2004). As found in Chapter 2 & 4, ambiguity is harder to detect than making inferences from an unambiguous situation and that spatial situations are harder to reason with than non-spatial situations. Furthermore, we found situational ambiguity detection is improved by probing the internal part of a participants constructed model in comparison to the external part. We suggest that situation model reading is deterministic in nature and that a specific question style influences reading behaviour and may encourage model revision. We also suggest that by discouraging the making of inferences based on a single premise, as well as discouraging making inferences from a single preferred model is a significant boon to ambiguity detection. However, in contrast to predictions we found that probing the internal part of the model does not benefit older and younger readers equally, and significantly improved older participant’s ability to make inferences about spatial situations. We suggest that older and younger participants make use of different strategies and that probing the middle of a model aids the older readers’ strategy more than that of younger readers. However, we also acknowledge that the experiment is not well geared towards finding out the specific strategies used but that this should be the focus of study in future. Lastly we found few significant differences as to which premise a model was built upon. Prior work investigating mental models has already established how a preferred model is generated and we again acknowledge that our experiment is not well geared to the detection of model preference.

Chapter 5.5: Preferred solutions

This thesis thus far has focused mainly on how, when and if readers of ambiguously described situations can detect this ambiguity by stating “there is not enough information”. As we have seen, readers are not particularly adept at this task. While readers are of course capable of detecting ambiguity as well answering questions about unambiguous situations, this thesis has not yet dealt with incorrect answers to ambiguous situations. The questions relating to which of the viable solutions is “prefered” when incorrectly answering questions about ambiguous situations is the focus of this subchapter.

General introduction

As discussed earlier in this thesis certain mental models are preferred over others for a number of different reasons. Simply put, readers have a tendency to create an order of objects/subjects from underspecified premises in a way that:

1. Creates a parsimonious order of objects/subjects while still obeying all the premises (Goodwin and Johnson-Laird, 2005). People tend to prefer mental models that are simpler and require fewer assumptions, as they are easier to understand and remember.
2. Creates a coherent order, i.e. one which does not violate a person’s prior knowledge of the world (Goodwin and Johnson-Laird, 2005).
   **Note: See chapter 1 page 5 for an example of such a violation*
3. Creates an order that is useful for the situation, in our case answering a comprehension question (and detecting ambiguity), (Nejasmic et al., 2015).

Despite these preferences being established it is likely that these vary depending on individual differences, such as prior knowledge or cognitive abilities, as well as situational factors such as time constraints or complexity of the task. More importantly it may be the case that the 3rd point, a preference to create a model that is useful for the current task is a main cause for readers of situational ambiguities to not detect this ambiguity and rather solve a model in a particular way. This “task dependency” is further explored in chapter 8.

What has not yet been explored in this thesis however are model preferences, i.e. which order is preferred when ambiguity is not detected. In this chapter we used the data from Chapter 5 in order to identify which order of objects/subjects is preferred (if any) when participants do not detect ambiguity. Point 1 above, “readers create a parsimonious model and prefer mental models that are simpler and require fewer assumptions” (Goodwin and Johnson-Laird, 2005), does not allow us to estimate which model may or may not be chosen. The models created from the premises throughout this thesis have the smallest possible number of components (three) with two underspecified components and therefore the two viable models are equal in their simplicity and numbers of assumptions. For example:

“The glass is to the left of the cup”

“The plate is to the right of the glass”

Creates two viable models:

“Glass…Cup…Plate”

And

“Glass…Plate...Cup”

Both the first and second model are created with one assumption; Glass, Cup, Plate, assumes that the glass and the cup are directly adjacent to one another, as ordained by premise 1. Conversely the second model assumes adjacency of the glass and the plate which is ordained by premise 2. Both of these models are coherent models, obey both premises, and do not violate any known laws of how the world works. Furthermore, both models are equally useful for solving the task of being able to answer a question about the model. It is therefore difficult to determine which of the two models would be preferred, if any.

In order to test this we first have to test whether or not any viable option was preferred over the other.

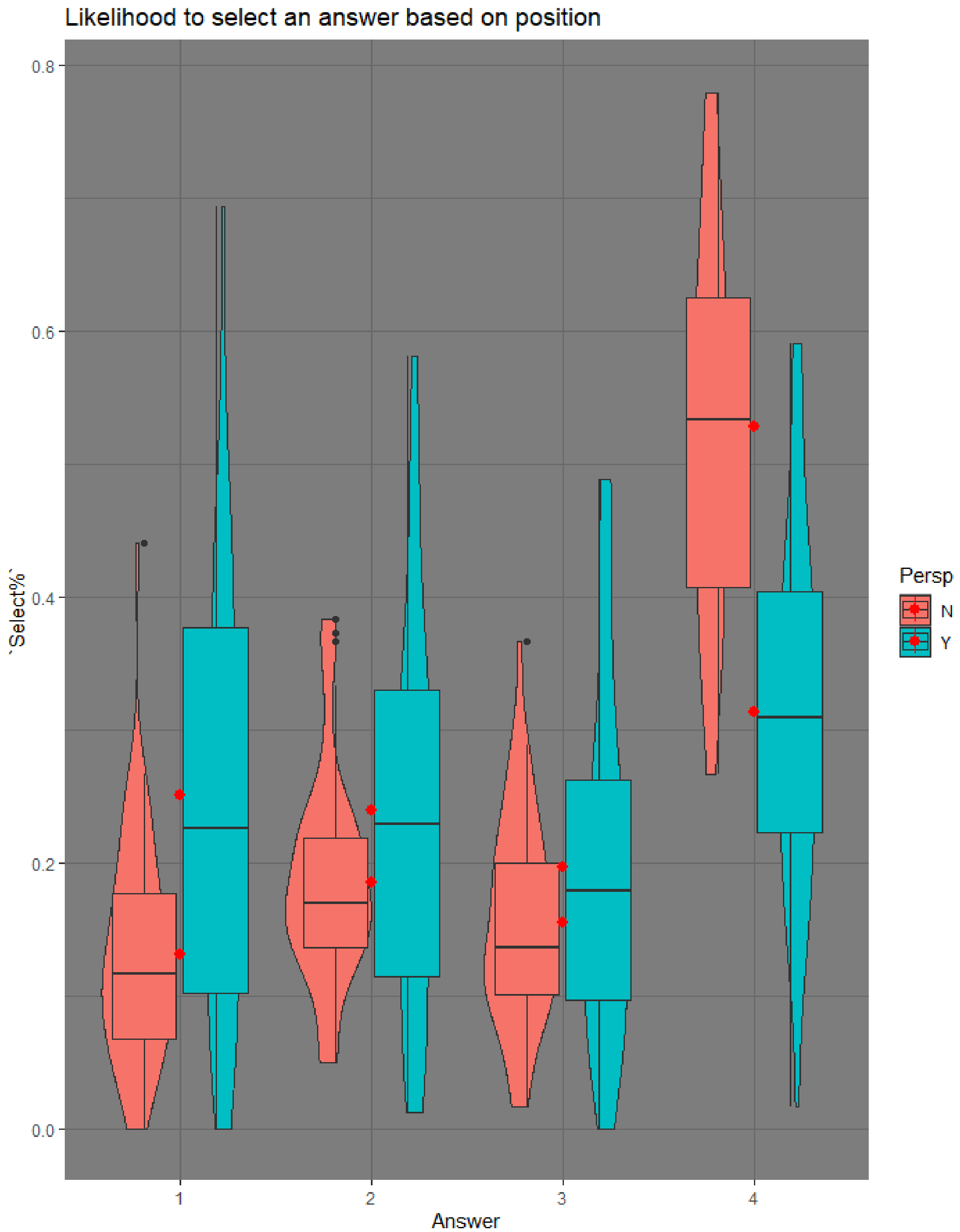
Analysis:

We used the data from Chapter 5 experiment 1 & 2, 254 participants both young and old (Mean age 48.91, SD=7.46, 188 Female) English speakers (readers) recruited online from Prolific and undergraduate students at Bournemouth university. We explored and analysed a number of possible influences on selection likelihood but found that only the content of Premise one appeared to influence model preference significantly. We analysed only questions about ambiguous stimuli where the participants chose an incorrect answer (ambiguity not detected).

Selection likelihood based on position

We investigated the selection likelihood of any given answer by its position on screen. We did this as a preliminary investigation to find out if any position was selected significantly more often based on its position alone. We found that, as expected, answer position 4 was selected the most often as this was the position that always contained “there is not enough information” which was the correct answer for questions about ambiguous stimuli. The details of this brief investigation are shown in Figure 1 below.

Figure 1.
Violin plot showing selection likelihood based on position (X= Position 1,2,3,4) between Persp (N&Y, the internal external manipulation from chapter 6) as a comparison between the two experiments.



Note: Raw data for this analysis can be found on the OSF for this thesis in the folder "Prefered solutions"

We removed all the instances where participants managed to correctly detect ambiguity (removing answer 4 from the data) and compared the selection likelihoods of incorrect answers only based only on their position on screen. Using a linear model we found no significant differences between the answer options as we can see in table 1 below.

Table 1.
A linear model showing selection likelihood by answer (position) comparing positions 2 and 3 to position 1.

Fixed effects	**Selection likelihood (%)**				
	b	**SE**	**df**	**t value**	**Pr(>|t|)**
Intercept	0.250	0.021	-	11.54	**<.001**
Position 2	-0.011	0.030	-	-0.37	0.71
Position 3	-0.053	0.030	-	-1.742	0.083

From Figure 1 and Table 1 it is possible to determine that it is unlikely that the position of the answer on screen played any significant influence on the preference on any particular model. However, it is possible to see that Position 3 is selected somewhat less often (though not significantly). This is likely due to the fact that the "anchor" answer, which was never the correct answer and is not a viable option for questions about stimuli, was in position 3 more often than in any other position.

Selection likelihood determined by premise

As mentioned earlier in the example of the glass, plate and cup both premises are equally simple and have an equal number of assumptions. Despite this we tested whether either premise 1 or premise 2 was preferred overall. In this investigation we removed the answers where participants were “correct” and selected “there is not enough information. We are therefore investigating end model preference, i.e. the order of objects subjects most preferred by participants. Irrespective of on screen location (Position one, two or three) one answer always assumed direct adjacency based on the first premise while the other assumed adjacency based on the second. We assume here that a preference for one or the other premise and subsequent model results in the other premise being considered “less true” and is “added” later, following the principle that mental models are constructed “bit by bit” (Johnson-Laird & Byrne, 1991). Table 2 below shows a Linear mixed model for selection likelihood based on premise (One ,Two or anchor) and the internal/external manipulation from chapter 6. Here we see that, while there is no significant difference in the selection likelihoods based on the internal/external manipulation, premise 2 was selected significantly more often than the anchor and premise 1. This is further shown in figure 2.

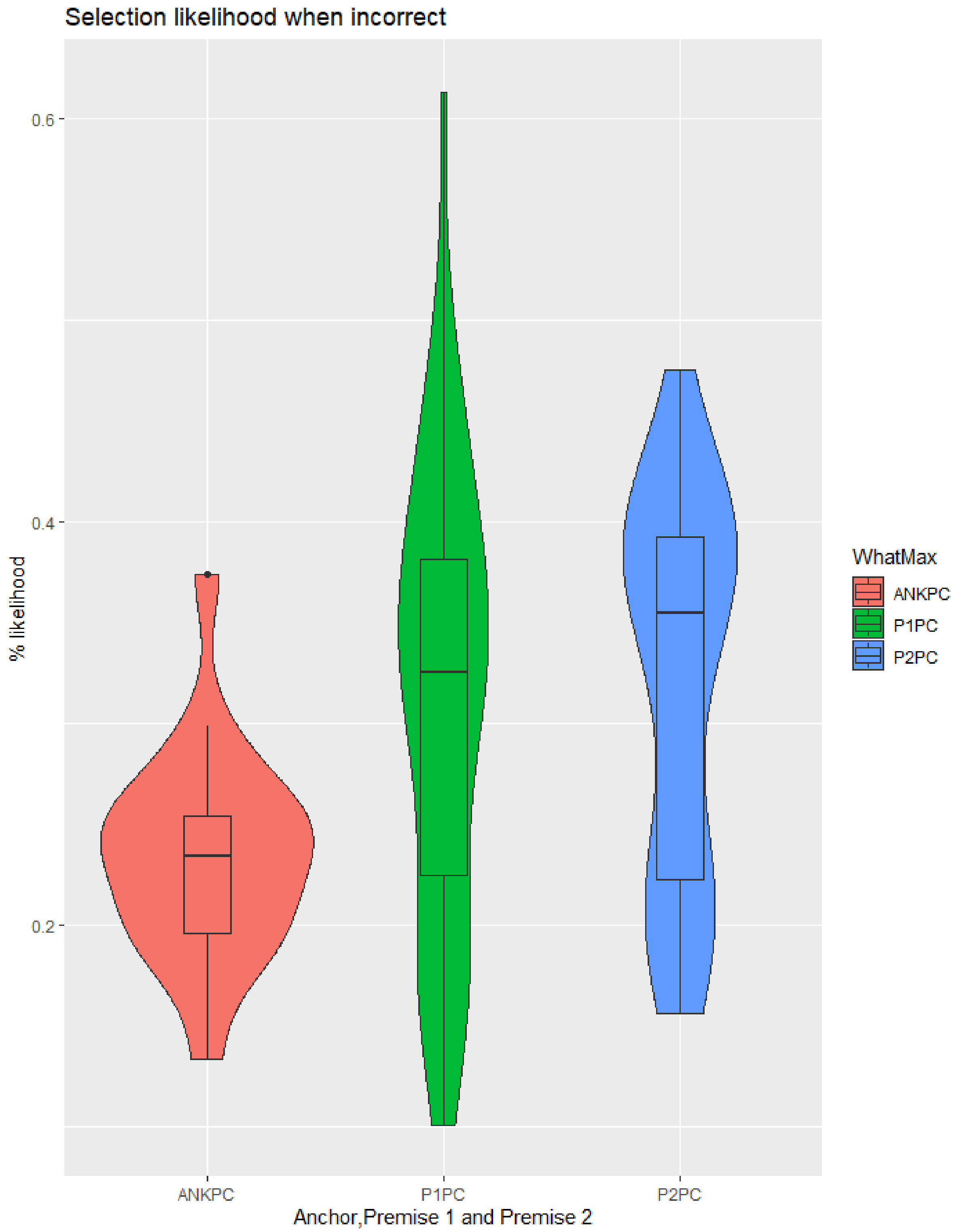
Table 2.

Selection likelihood based on premise (comparison to the anchor) by internal/external.Random effects for this model include the likelihood for the answer containing premise's 1,2 and the anchor to occur any of the positions (1,2 or 3).

Fixed effects	**Selection likelihood**				
	b	**SE**	**df**	**t value**	**Pr(>|t|)**
Intercept	0.322	0.014	28.77	22.153	**<.001**
Premise1	0.018	0.025	86.014	0.732	.466
Premise 2	-0.063	0.017	87.57	-3.575	**<.001**
External	-0.046	0.014	86.83	-3.126	**.002**
Premise 1 :External	<.001	0.025	85.94	-0.033	.973
Premise 2 :External	0.0170	0.015	86.34	1.089	.279

Note: Despite answer position not playing a significant role in selection likelihood, there may have been a small discrepancy of selection likelihood based on which premise's answer was located in which answer position as this was not balanced (e.g. anchor answers often but not always appeared in position 3). Because of this we ran a weighted version of the same model where this imbalance was accounted for by adjusting the weights to assume each premise was in each answer position 33% of the time. This analysis found that premise 2 was still selected significantly more often than premise 1 or the anchor. This analysis can be found in the OSF document for this thesis in the folder "Preferred solutions".

Figure 2.
Violin plot showing the predicted likelihood to select Premise 1(P1PC), Premise 2 (P2PC) or the Anchor (ANKPC) when participants failed to detect ambiguity (combined internal external conditions).



As can be seen from Table 2, the effects of the internal/external manipulation appear to “dull” the differences between selection likelihood. As we have learned from chapter 6 this may be due to the internal condition making it easier for participants to detect ambiguity and as such the selection likelihood better reflect random chance selection than preference by participants. Furthermore, the chance to select the anchor answer may also reflect this as the anchor is not and cannot be the correct answer in any of the situations. Therefore we can take a closer look at the above described effects by:

1. Removing the results from the internal condition (leaving the results of 106 participants).

And

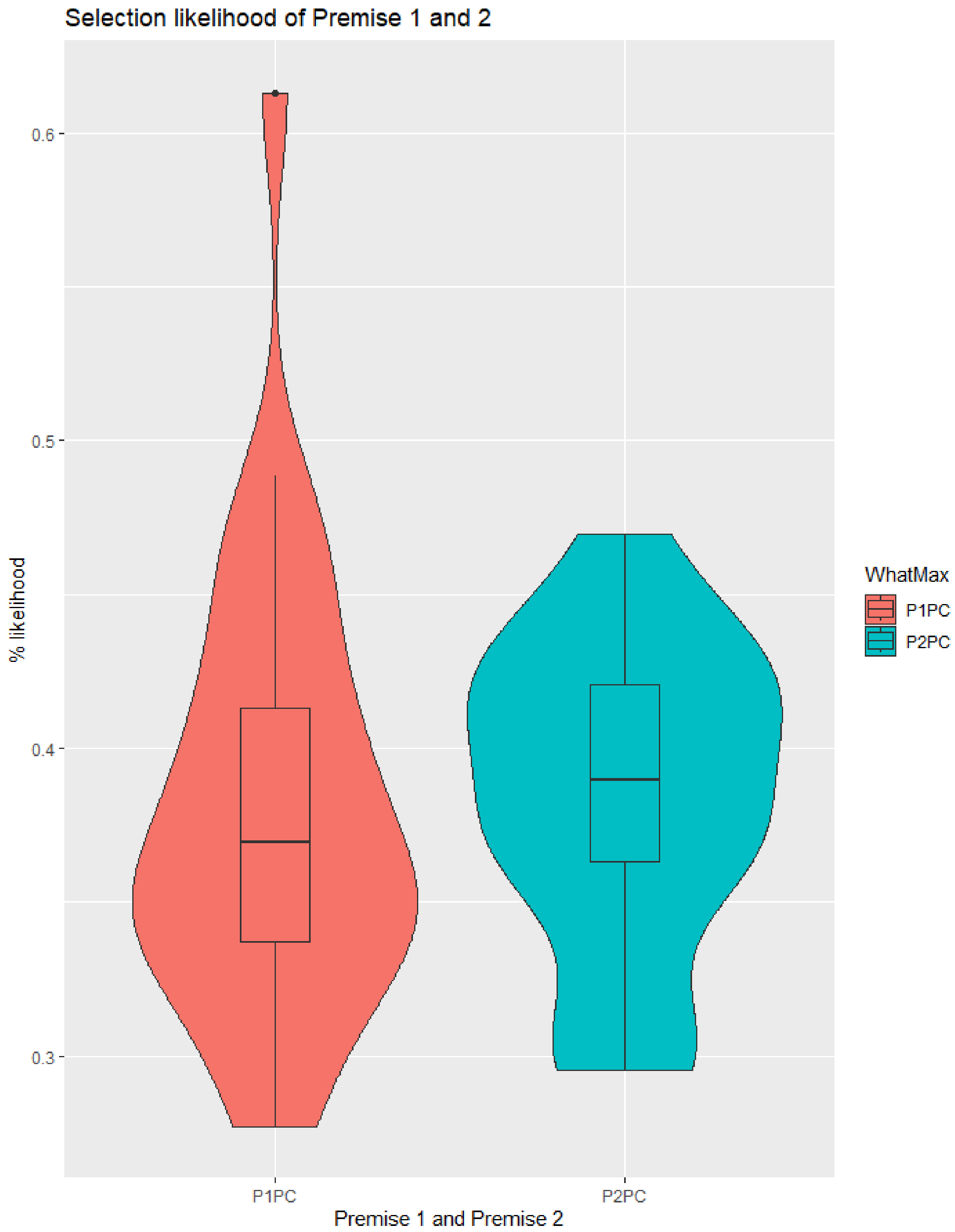
2. Removing the selection likelihood of selecting the anchor.

Using the same main and random effects in a new linear mixed model we found that the answer based on premise 2 was selected significantly more often than that of premise 1. This is shown below in table 3 and visualised in figure 3.

Table 3.
Selection likelihoods by answers based on premise 1 or premise 2.

Fixed effects	**Selection likelihood**				
	b	**SE**	**df**	**t value**	**Pr(>|t|)**
Intercept	0.340	0.018	9.810	18.311	**<.001**
Premise2	0.121	0.039	42.453	3.098	**<.001**

Figure 3.
Violin plot showing the selection likelihoods of Premise 1 and Premise 2 coded answers.



Discussion

In this subchapter we examined the tendency for model preference. It is likely ,at least in the case of our stimuli, that the position of the answer does not play a significant role in the preference for an answer (see Figure 1 and Table 1). However, as can be seen from tables 2 and 3, as well as, figures 2 and 3 that there is a significant tendency for participants to select an answer which is based on premise 2 rather than premise 1 or the anchor.

However, this may not be as simple as it first appears. The answer does not directly refer to any particular premise at all, rather the position of a single object/subject described in the premise. The way in which we interpret this “tendency” for participants to select premise 2 relies on the assumption that the participants have read and understand premise 1. For example:

Premise 1 describes:

“The glass is to the left of the cup”

Premise 2 describes:

“The plate is to the right of the glass”

And the following question asks a reader

“ Which object is farthest to the right?”

A: The cup

B: The glass

C: The plate

D.There is not enough information

We can firstly “eliminate” the glass from the possible solutions to this as it, being the furthest left and being the object used to describe the location of the other two, is the anchor. We are now left with the ambiguous part of the situation where the direct relationship between **the cup** and **the plate** is unknown. Let us imagine that we have not “detected” that the relationship between the two is ambiguous, we now have to select either **the cup** or **the plate** as our answer. Both the left over viable options are based on models which can be used to glean an answer to the question and have the same numbers of assumptions. Option A, **the cup**, assumes that **the plate** is directly adjacent to **the glass**, which we could assume using premise 2 (*“The plate is to the right of the glass”).* Option C, **the plate**, on the other hand assumes that **the cup** is directly adjacent to **the glass** using premise 1 (*“The glass is to the left of the cup”)*.

Using what we found in the analyses above it is possible to determine that participants are more likely to select an answer based on premise 2 but described in premise 1 i.e “**The plate** is directly to the right of **the glass** (“plate right of glass”, premise 2) and **the glass** is to the left of **the cup** (premise 1). Therefore the cup is furthest right as it isn’t directly adjacent to the glass.” It is possible that as the two premises are “constructed” into a model by the reader one premise “takes priority” over the other when adjacency to the anchor is considered while the other is “slotted into” the model so long as this does not violate the preferred premise. While this is somewhat convoluted it does follow the principle that mental models are constructed “bit by bit” (Johnson-Laird & Byrne, 1991). Though how and why this preference is determined by readers is as of yet not known and above and beyond the scope of this thesis.

It has to be noted that the logic that may dictate premise preference can also be reversed i.e.“**The glass** is directly to the left of **the cup** (premise 1) and **the plate** is to the right of **the glass** (premise 2). Therefore **the plate** is furthest right as it isn’t directly adjacent to **the glass**.”. This would result in the selection of the other answer which while significantly less likely, did happen. The premise preference of participants is therefore clearly not universal. One possible explanation is that people use their prior knowledge and assumptions about how objects are typically arranged to fill in the gaps in the underspecified order. For example, if people typically place forks to the left of knives (as is good and proper) they may assume that this is the intended order in the absence of additional information. Additionally, people may rely on their perceptual biases, such as the tendency to perceive objects in a left-to-right order, to make sense of the arrangement.

Given that the stimuli covered 48 spatial and non-spatial situations with various degrees of potential assumptions and perceptual biases it is possible then that these influenced the model readers created from which to answer the question. In an ideal situation we could create stimuli that are devoid of these biases in an attempt to see whether or not a reader's preference for a premise 2 based answer is the "default" or perhaps some kind of primacy effect. However, even without such biases from the stimuli preference could be based on a number of factors, including prior knowledge and experience, personal biases, and cognitive processes such as mental simulation and inference. The degree to which people are able to resolve ambiguity and arrive at a consensus interpretation can vary depending on a number of factors.

What this short investigation does show us (as well as the results of chapter 6) is that ambiguity detection and the selection of a particular model when ambiguity is not detected are the result of readers creating a model that is suitable for the task. It is not yet known if a model built to create a clear order of objects/subjects is suitable for use in ambiguity detection (though unlikely) or if a model created for ambiguity detection is suitable for model solving. This question is addressed in Chapter 5 and 6.

Chapter 6: Unambiguous answers from ambiguous situations: Task dependency in situational ambiguity reading

In keeping with open science, the full reproducible analysis code, raw data and supplementary materials are available on the OSF link provided below.
(Make the OSF for this)

Abstract

Ambiguity detection from a situation described in a text, has been relatively absent from research in mental model research and the strategies involved are not well understood. When made up of two premises, ambiguous models always contain one Anchor object/subject and two objects/subjects in an ambiguous relation to one another. In this study we presented participants with a series of paragraphs of stimulus text, describing a situation, in two blocks probing different parts of the situation. We found significant changes, in terms of comprehension accuracy, reading/reasoning speed and encoding behaviour, when participants who were (self) “taught” how to locate an unambiguous anchor of a model switched to having to detect ambiguity. We found that “teaching” participants to detect ambiguity in the first block of the study made it significantly easier for them to locate an anchor in the second block and that switching from identifying the anchor to detecting ambiguity was significantly harder. We suggest that this is due to at least two different strategies being employed by participants. One, more holistic and cautious approach from ambiguity detection, and one focussed and less cautious from unambiguous model part identification. We also replicate prior findings that situational models describing a spatial situation are harder to comprehend than non-spatial situational models. Lastly, we expand upon the prior findings that ambiguous situational models are harder to comprehend than unambiguous ones as we found that questions probing ambiguous model parts are also harder to answer than unambiguous ones.

Reflections on Chapter 5:

Older readers reread more than younger readers (Rayner et.al,. 2006), in Chapter 4 we found that experiment 1, older and younger readers seemed to re-read the same amount but that in experiment 2 older readers reread significantly more often than younger ones. With this it is possible to conceive of a situation whereby the "risky" reading strategy is beneficial to older readers by facilitating a greater amount of re-reading. However, as re-reading is beneficial to both young and old, younger readers also adopt this strategy in order to better comprehend the model. Furthermore, when ambiguity is more easily detectable, as appears to be the case in experiment 2, we find that re-reading is increased significantly in older readers compared to younger ones. It therefore seems to be the case that older readers modulate their "risky" reading strategy to increase their rate of re-reading. We also found specific target word re-reading increases with age (chapter 6). If we apply the findings about older readers eye-movements of chapter 6 to the task dependency findings of chapter 4 it is possible to deduce that ambiguity detection and model solving is high dependent on encoding strategy, which is dependent on eye-movement/reading strategy. Eye-movements and reading strategy are therefore task dependent, as well as age dependent. It is also highly likely that working memory capacity plays a key role in the encoding and processing of information (Blott et al., 2021; Nicenboim et al. 2015; Schurer, et.al t,. 2020; Schwering and MacDonald, 2020), though how this influences eye movements and the interpretation of situational ambiguities is not yet known and unfortunately not a key method of investigation in this thesis and works contained therein.

One way in which it could be possible to establish working memory differences, eye-movements and situational comprehension comes in the form of task switching. Assuming that eye-movements are in fact modulated by the task at hand it would be possible to see how a reader's behaviour changes when the task changes. It has been shown that task switching and working memory are interconnected (Liefooghe, et.al,. 2008) and that there is an inherent cost to working memory when task switching occurs (Vandierendonck, 2021). Moreover, working memory plays a key role in switching attention to specific elements, depending on the task (Unsworth and Engle, 2008). Therefore, it is highly likely that task switching ability may give additional insight into a person's working memory ability and the role this plays in the detection of ambiguity in situational models. Additionally it is known that working memory declines as we age (Bopp & Verhaeghen, 2007; Borella et al., 2008; De Beni et al., 2004; Fiore et al., 2012 ; Park et al., 2002) and therefore the ability for an older person to task switch may be significantly different than that of a younger person. If we reframe the risky reading strategy of older readers as a way of mitigating a lower working memory capacity it makes sense to suggest that older readers may be better suited to higher working memory capacity tasks. This is evidenced in chapters 4 and 6 where we find fewer differences in terms of accuracy. However, in those studies the tasks required of the reader switched regularly, from unambiguous model solving to ambiguity detection and back again within a few items. As we suggest that ambiguity detection is highly task dependent, and encoding is modified depending on the task, we need to create a paradigm in which a task switch and its influences can be studied. If ambiguity detection is truly task dependent we would see a different encoding strategy.

In chapter 6, we do not see much of a difference in regression rate between ambiguous and unambiguous stimuli. However, this does not mean that encoding strategy is not different. Instead it may simply suggest that a reader’s encoding strategy falls into a default encoding strategy, which is not specifically tailored to the presence or absence of ambiguity. As previously suggested ambiguity detection requires the reader to combine the two premises, something which can only be done once the second premise has been read. Therefore the reader cannot modulate their reading behaviour on information not yet available to them. However, what is more available to them is information on subject matter, spatial vs non -spatial, as this information would be available to the reader as of the first premise, and possibly even before. In chapter 4 we found that readers have significantly higher regression rates in the re-reading of target words in spatial stimuli. If we combine this with the overall lower comprehension accuracy of spatial questions then it may be the case that readers have a tendency to modulate their reading behaviour to compensate for the additional difficulty presented by spatial situations. However, they are only able to do this as it is relatively clear as to what type of situation is described. Something that cannot be done for ambiguous situations. In short the reader uses context specific information in order to modulate their reading behaviour.

Would it be possible to create a stimulus text in which contextual information assists the reader in their ability to answer comprehension questions about ambiguous stimuli? In the following chapter we create an experiment in which every stimulus is ambiguous, and the question probes different, ambiguous or unambiguous aspects of this model.

1.0 Introduction:

Thus far in this thesis we have found that readers are able to detect situational ambiguity, but they are not very good at doing so. However, we have also repeatedly found no indication that reading an ambiguous situation takes no longer than reading an unambiguous one. While this might indicate that the encoding of the premise information is the same for ambiguous and unambiguous situations, we also found that the re-reading of premises is often significantly higher in ambiguous situations but only when probing the internal part of the model. It therefore stands to reason that probing the internal part of the model influences the way in which premise information is encoded and, judging by significantly higher comprehension accuracy, processed. In this chapter we explore if probing the middle part of a model, in contrast to an always unambiguous external part of the model, may encourage better model encoding and processing.

While it is hard to know which specific strategy, as there are a variety (Schaeken et al 2000), is used in the processing of a situation, we are able to discourage certain strategies not conducive to ambiguity detection. Chapter 5 established that encouraging a more holistic understanding of the situation led to higher ambiguity detection in both older and younger participants. It seems that probing the internal part of the model, rather than the external part, leads to a different encoding and reasoning strategy. It stands to reason that, should ambiguity detection be deterministic, the question style leads to a more holistic construction of a model even if probing an unambiguous part of an ambiguous model later.

In Chapter 5 we speculated that participants use a preferred model or may create a "lazy model" based on only a single premise where the location of the third object/subject is inferred but that this is not useful for detecting ambiguity. By inhibiting a participant's ability to use a lazy model and encouraging model revision (leading to higher rates of realising that there is more than one viable model), we managed to increase ambiguity detection. It has to be noted at this point that while investigation into strategies employed to construct and reason with mental models is extensive (e.g Franz & Garnham, 1990; Baddeley, 1998; Johnson-Laird & Byrne, 1991; Rauh et al., 2005; Goodwin & Johnson-Laird, 2005; Knauff & May, 2006), little is known about how these strategies may differ when a "new task" is presented.

In this study we "taught" participants to use a reasoning strategy which allowed them to better understand the situation as a whole, to better make inferences about an unambiguous part of a model. We did this by using only ambiguous situations and probing the internal part of the model for the first half of the study. The questions then switched to probe the location of the anchor, which is always in an unambiguous position even in ambiguous situations. We contrasted the "internal first" participants to "anchor first" participants who were "taught" using questions probing the anchor for the first half of the study and then switched to probing the internal part of the model. We predicted that encouraging participants to use a strategy which results in a more holistic comprehension of an ambiguous situation is beneficial even when not detecting ambiguity, and that encouraging a non-holistic strategy for locating the anchor of a model is no longer beneficial when detecting ambiguity.

For example, if we consider the spatial relationship between three objects ***A***,***B*** and ***C***:

Example 1:

*"**A** is left of **B**"*

and

*"**C** is right of **A**"*

As we have learned in the previous chapters, it is impossible to discern the spatial relationship between B and C as the situation is ambiguous. However, the location of A is quite easily discernible as the object farthest to the left, which is what makes "***A***" the anchor. It is important to note that mental models are constructed "bit-by-bit" (Johnson-Laird & Byrne, 1991) and participants "check" that their interpretation is consistent with each of the premises, models which defy one or more of the premises are generally not preferred (Jahn et al ., 2007). This construction process leads to a single viable model from which the location of A can be gleaned relatively easily in both an ambiguous and unambiguous situation. Alternatively the location of "***A***" can be inferred using a lazy model, using either one of the premises alone. Both strategies create models which are useful for locating "***A***" (Nejasmic et al., 2015). Regardless of constructing a single viable model or using a lazy model by continuously probing the location of the anchor, in the case of the example object "***A***", a participant may be "taught" that considering a single model is all that is required in order to comprehend the model. However, this strategy ignores any other model that holds true with the premises and so this strategy is not conducive to ambiguity detection. On the other hand if the internal part of the ambiguous situation is probed. In the case of Example 1, if ***B*** or ***C*** are in the middle. Participants may start to develop a more "cautious" approach, checking for other consistent solutions. We predict that this approach takes longer than one which is less "cautious" where participants have been probed on the non-ambiguous anchor.

The task, locating the anchor, appears easier and re-evaluation may seem to be a waste of time.

1.1 This study

This experiment focuses on how ambiguous situational models are interpreted based on which part of the model is probed by the question. Chapter 5 found that ambiguity detection in situational models may be deterministic. In short we found that readers are better at detecting ambiguity when questions probe the internal part of the model than readers attempting to identify the position objects/subjects on the periphery of the model. We speculate that by “teaching” participants to be more cautious when evaluating/reevaluating the situation by discouraging the use of a preferred model or a lazy model, we can improve model comprehension as a whole. Conversely by “teaching” participants, using anchor probing, that a single preferred model or a lazy model is good enough, we can cause difficulties when participants switch to detecting ambiguity. Our study utilises counterbalanced blocks where the first half of the questions probe either the ambiguous aspect of the model, or the non ambiguous aspect of the model followed by the other in the second half. We predict that while all participants will experience some difficulty when the question probe is switched, leading to higher SRT,QRT and lower accuracy, the group which was “taught” using the internal probes will perform significantly better in the second half than the group which was “taught” using anchor probes. However, we also investigate how the “encoding” behaviour may change based on which part of the model is probed. We also investigate line by line reading and re-reading as a factor of which question style participants were “taught” with. Here we predict that the increased need for a holistic model in order to detect ambiguity, leads to higher rates of re-reading when the stimulus text is presented but only when participants are “taught” using the internal probe. We also speculate that the reverse may be true for when participants are “taught” using the anchor probe.

In this study, participants were presented with a situationally ambiguous stimulus describing a spatial or non spatial situation. A simplified example of this would be:

Example 2.0:

*"**Toby** was slower than **Harry**.*

***Harry** was faster than **Suzie**"*

In this situation, the relationship (in terms of speed) between Suzie and Toby is unclear but it is certain that out of the three people Toby is the fastest. Half of the participants would begin the study answering questions about the "Anchor" object/subject, in this case Toby, and be asked to identify him. For example:

Example 2.1:

"Which of the three is the fastest ?"

However, the other participants would answer a question about the ambiguous element of the model, in our case Harry and Suzie, and be tasked with identifying ambiguity. For example:

Example 2.2:

"Which of the three is neither the fastest nor the slowest?"

In both the anchor and internal probe the answer options are the same and would consist of; "***Harry***, ***Suzie***, ***Toby*** and ***There is not enough information***". In the case of the question in Example 2.2 both ***Harry*** and ***Suzie*** are viable but participants are informed that should more than one model be viable "***There is not enough information***" is the correct answer. At the halfway point of the experiment (24 viewed stimuli) the participants' questions would switch question types. By blocking the questions in such a way it was possible to investigate the associated "cost" when switching tasks from identifying ambiguity to anchor identification and anchor to ambiguity identification.In the case of the anchor questions, there would be enough information to answer but not in the ambiguity probing questions.

Lastly, as in Chapters 2,4 and 6, we investigated potential differences between spatial and non-spatial models. Specifically we wanted to deepen our understanding as to why participants found it significantly harder to answer questions about both ambiguous and unambiguous spatial models compared to non-spatial ones. For example, participants performed significantly worse in identifying the ambiguity in the situation:

Example 3.0:

" There was a cup to the left of the laptop.

"There was a pen to the right of the cup"

Compared to:

Example 3.1:

"Sally was weaker than Suzan.

Tia was stronger than Sally"

Despite both of the models creating the "same" model in terms of order. Participants also performed significantly worse in answering questions about unambiguous spatial models.Our initial research attributed this difference to greater semantic interference in spatial situations which is known to cause problems in mental model processing (Knauff & Johnson-Laird 2002;Knauff & May, 2006). However, we also outlined other possibilities such as problems with scene visualisation (Ford, 1995; Bacon et al., 2008) which make spatial situations harder to reason with than non-spatial situations. When recalling the situation irrelevant descriptives may interfere with encoding and recall of the premises (Dewar et al. 2007; Craig and Dewar, 2018), which may not happen with non spatial, more abstract, concepts such as wealth or popularity.We also outlined that the difficulty in spatial comprehension may come from complexities relating to the axis on which a situational model can be reasoned out. In short, we suggested that the difficulty in solving/identifying ambiguity in spatial models came from the “enforced” axis (e.g left to right) which may not have been the reasoner’s preferred reasoning axis. This axis plays less of a role when constructing non-spatial models.

This study displays spatial and non-spatial stimuli in random order within both blocks and so these are not part of the question switching paradigm, though this may be useful for future research. The spatial vs non-spatial manipulation of this study is therefore both a confirmatory and exploratory manipulation whereby we aim to expand upon our prior research into spatial/non-spatial text interpretation and also investigate this through anchor/ambiguity question switching manipulations. We predict that the differences in comprehension accuracy found across this thesis will remain in this study as we are not manipulating the spatial and non-spatial stimuli.

Lastly, as in Chapter 5 this study used text masking to allow us to approximate how long participants spent reading a particular line of text. AsIn lieu of eye-tracking using this mask made it possible to glean information about how premise information is read and re-read. Further, details about the text masking can be found in the Materials and design section. We predict that the line by line reading time will follow a similar pattern as found in Chapter 5 whereby premises were re-read more often than non-premise lines.

1.2 Participants

We collected data from 80 native English speakers (Mean age 35.1, SD=12.1, 45 Female, 1 other) recruited online from Prolific.co and undergraduate students at Bournemouth university. 69 of the participants were recruited via prolific while 11 were]undergraduate students from Bournemouth university. All participants were naïve as to the ambiguity manipulation. The study was approved by the Bournemouth University Research Ethics Committee (ID 27563). Each participant was informed of the experimental procedure and provided informed electronic consent. Due to potential cultural biases (Andrews et al., 2013; Maass et al., 2007) the study was limited to native English speakers living in the UK. Participants received monetary compensation (£7.44 per hour) or university course credit regardless of how they performed in the study.

1.3 Materials and Design

The experiment had a 2x2 mixed design, non-Spatial vs spatial within subjects and block order (Anchor question first vs Ambiguity question first) between subjects. The trials were presented in pseudo-randomised order and the trial texts described an ambiguous situation. This line of enquiry is exploratory and we did not perform a power analysis in order to estimate power required to detect the effects of the trial blocks.

The experiment consisted of 48 trials, in each trial participants were presented with a paragraph describing a situation. 24 described a spatial situation and 24 described a non-Spatial one situation, in contrast to prior chapters all situations were ambiguous. Each paragraph described an ambiguous situational model. Each stimulus paragraph was followed by two questions, shown one after another. The initial question probed the model, either through its anchor or its ambiguous element, depending on the participant counterbalancing. The second question probed general comprehension of the text. Below we illustrate an example stimulus in text form:

Example 4:

"The town of Rockport hosted three major events that always drew in large amounts of crowds. ***The folk music festival was more popular than the Rockport Grand Prix****. Both were important sources of income in the town.* ***The Sailing festival was more popular than the Rockport Grand Prix****, but this didn't matter to the townsfolk. Rockport was one of the most up and coming towns in the area. The yearly events were beginning to draw in crowds from further and further afield."*

The stimulus above describes an ambiguous non-spatial situation, the premises (**bolded** in the text) create a situation where it is not possible to tell which festival is the most popular. Another example, this time for a spatial stimulus, is shown below:

Example 5:

"Spice Island lay just three miles offshore. The island didn't have a lot to offer in terms of comfort but there were a handful of people that called it home. ***There was a small weather station to the left of the small dock used to load and unload supplies. There was also a supply store to the left of the docks.*** *It was overpriced because everything had to be imported from the mainland. Many islanders didn't mind spending more on supplies if it meant avoiding a long trip to the mainland."*

<u>Note:</u> *In the experiment participants would not see the full text, only one line at a time, an example of this is shown in Figure 1 in the procedure section.*

The premises, all describing an ambiguous situation, were created by relating **A** to **B** (e.g "The phone to the right of the cup") and **C** to **A** (e.g "The pen to the left of the phone") or **C** to **B** (e.g "The pen to the left of the cup") in two separate sentences. These relation are counterbalanced whereby half of the second premises relate **C** to **A**, and the other **C** to **B**

The “distance” between the premises (the amount of text between the premises) was counterbalanced (mean gap= 1.54 sentences, SD= 0.45 sentences) , the examples above show a 1 sentence and 0 sentence gap respectively. A full list of stimuli and questions can be found on the OSF for this experiment.

Readability analysis used the readability package for R with the Flesh-Kincaid measure (Kincaid, 1975) using the quanteda package (Benoit K, et.al 2018), while the word frequency analysis used SUBTLEX-UK (Keuleers, Lacey, Rastle and Brysbaert, 2011).Table 1 shows the mean readability statistics for the stimuli and questions.

Table 1:

Table showing mean readability metrics of the stimuli and questions.

(A full set of stimuli and questions can be found via the OSF link)

Stimuli

Condition	**Question Type**	**Word count**	**Word Length in letters**	**Readability**	
				Flesch-Kincaid	**Word Frequency (Zipf)**
Spatial	N/a	85.58 (4.07)	4.44 (0.19)	7.05 (1.62)	5.95 (1.32)
non-Spatial	N/a	85.50 (3.69)	4.41 (0.26)	7.17 (1.26)	5.99 (1.15)
			Questions		
Spatial	Anchor	6.04 (1.36)	4.17 (2.17)	3.96 (2.43)	6.16 (0.30)
Spatial	Ambiguity	8.58 (1.66)	4.17 (2.17)	4.14 (1.29)	6.19 (0.26)
non-Spatial	Anchor	6.08 (1.47)	4.42 (0.53)	3.87 (2.81)	6.34 (0.19)
non-Spatial	Ambiguity	9.75 (1.88)	4.43 (0.42)	4.36 (1.63)	6.21 (0.30)

The full list of stimuli and questions in their various conditions can be found through the OSF repository for this study. The questions did not significantly differ in their readability or word count though they did differ in which part of the model these probed. More on this difference can be seen in the procedure section below.

1.4 Procedure

Participants were provided details of study participation needed for providing informed consent. After participants consented to participate, they were given instructions on what the task involved, further, they were primed to answer questions which had no clear answer (due to the situational ambiguity manipulation) with the answer “There is not enough information”.

Participants saw a screen with the stimulus text masked by the letter “X” along with the instruction “Please press the down key to begin reading” (see Figure 1). Pressing the down key showed the next line of text and simultaneously remasked the prior line. Up key presses reversed this allowing readers to move back up through the text. The stimuli were always presented in 6 lines. The 7th line was a further instructional text (“please press space to continue”) which took participants to the questions.

To account for different size monitors, the stimuli size and position were defined in PsychoPy’s “height units which on a 16:10 aspect ratio screen are represented as -8,-5 as the bottom left and 8, 5 as the top right, meaning that the text appeared in the centre of the screen . All text was presented centred on the screen (Davenport & Smith, 1965) in white on a black background using the Consolas font with a height of 0.03 units. The entire experiment code can be found via the OSF link and works with all common computer screen aspect ratios. The experiment was hosted on Pavlovia.org.

Figure 1:
Stimulus as presented on screen with the first line of text unmasked

Participants could read at their own pace and could only advance if they pressed the spacebar when line 7 was unmasked. This line by line self-paced reading technique allowed us to calculate reading time for the individual premises, as each line contained a single sentence. The first five lines contained 15 words, while the last line (line 6) contained between 5 to 16 words. Two of the lines contained the premise information (1 each) did not differ significantly from non-premise lines in terms of word count (Premise lines: mean=14.98,SD=2.36; Non-premise lines: mean=15.10, SD= 2.45) or readability (Premise lines Flesh-Kincaid: mean=4.98, SD=2.47; Non-premise lines Flesh Kincaid: mean=5.06, SD=2.54). The number of lines between premise 1 and premise 2 was not accounted for during counterbalancing but did not differ significantly across conditions (mean= 2.40, SD= 1.05).

Two questions followed each stimulus. The first question asked about the spatial or non-Spatial relationship described in the premises and had 4 answer options (see Figure 2a & b). The question was displayed at X=0, Y=0.3 at 0.05 height units. The options were displayed at the bottom of the screen in white, equidistant, clickable boxes (Width units 0.38, Height units 0.2) at positions along the Y axis at -0.35 units and the X axis at -0.675, -0.225, 0.225, 0.675 units respectively.

Figure 2a:

Question screen showing an Anchor question

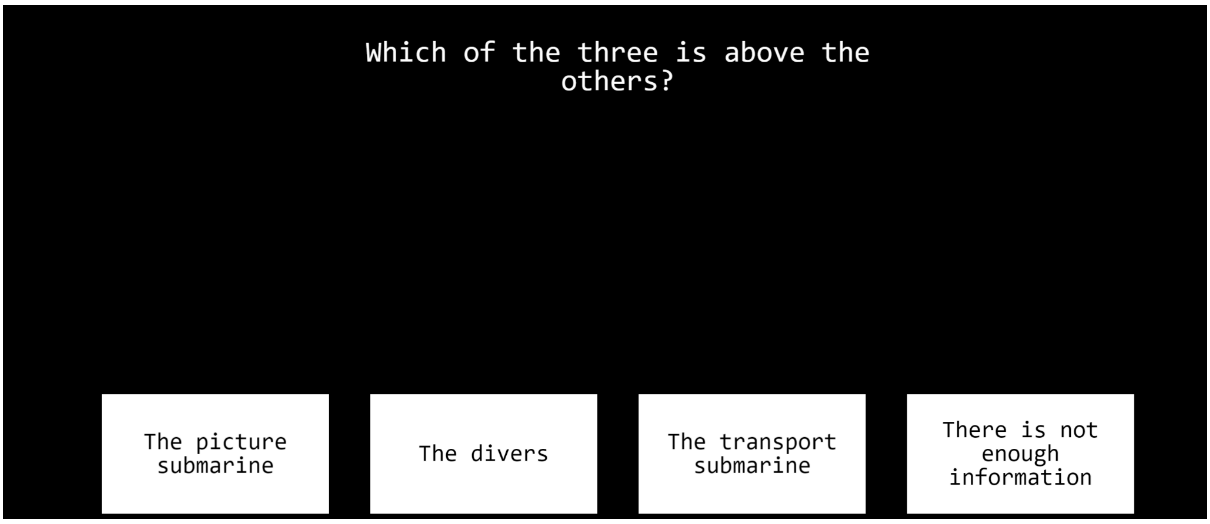
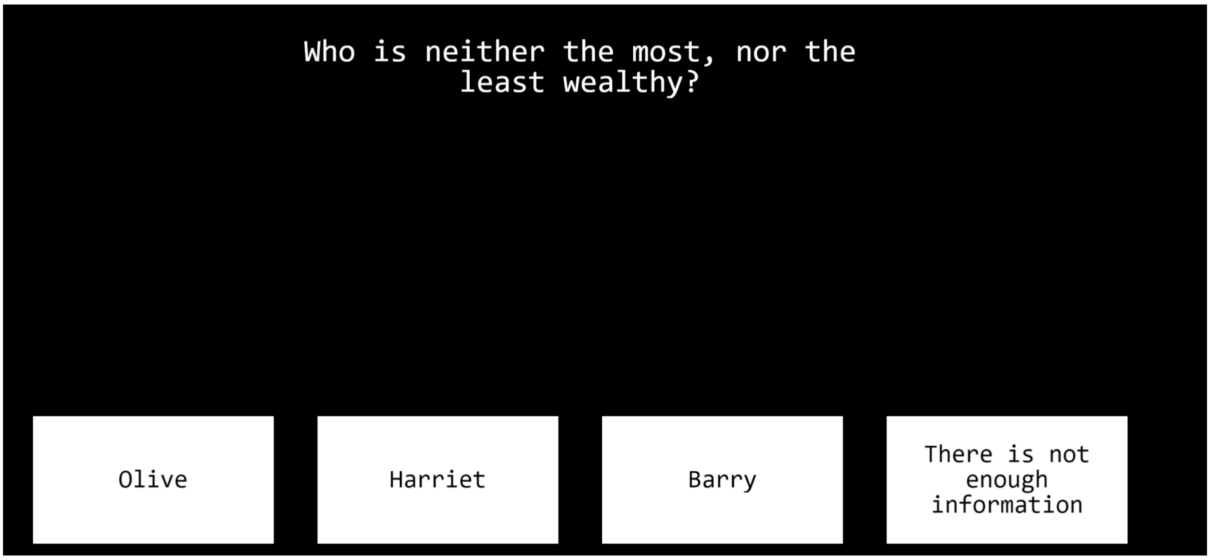


Figure 2b:

Question screen showing an Ambiguous Question



The four options always had two options that were equally plausible in the ambiguous condition, though only one of these would be correct in the anchor version. A third option was categorically incorrect and a fourth was "There is not enough information" which was always correct for the ambiguous question condition conditions. The questions and answer choices didn't differ between ambiguous and anchor versions of the questions. Participants would indicate their selection with a mouse click.

As can be seen in Figures 2a the Anchor questions probed the "extremes" of the model, based on the first premise in the stimulus, while the ambiguous question probed the "middle" of the model as seen in 2b. Participants completed the study in a single session. The experiment lasted about 25-45 minutes on average and participants could take short breaks when needed.

1.5 Analysis

The data were analysed using (Generalised) Linear Mixed Models using the lme4 package v.1.1-21 (Bates et al., 2015) as well as Generalised Additive Mixed Models (GAMMs)(Baayen et al., 2017; Sóskuthy, 2017; Wieling, 2018) from the mgcv package (Wood, S.N. 2017) in the R software v. 4.0.4 (R Core Team, 2021). We used contrast coding for stimulus type (ST) (Spatial 1, non-Spatial 0), question type (QT) (Ambiguous/Internal 0 , Unambiguous/Anchor 1), and block type (BT) (Ambiguous First 0, Unambiguous/Anchor First 1) in the analysis. Millisecond reading times were log transformed for total question, total stimulus reading times and line by line reading time. We included random effects of Item and Participants in all analysis, we are aware that more complex models can converge for specific analysis, however this allowed all models to converge with the same “base” random effects.

Modulation by trial order was tested using Generalised Additive Mixed Models, GAMMs function a lot like Generalised mixed models but predictors are specified as smooths. These smooths represent the weighted sum of a number of base functions (Baayen et al., 2017). In this model, gaussian process smooths were used as the base function for Sequence (across the experiment). Block type (BT) interacting with sequence within the block and Stimulus Type (ST) interacting with sequence within the block were set as factor smooths. Random slopes/intercepts were added for subject and item (The same as the GLMMs and LMMs). The GAMM models were fit with the “mgcv” v.1.8-26 R package (Wood, 2017) and visualised with the “itsadug” v.2.3 R package (van Rij et al., 2017). The remaining graphs were generated with ggplot2 (Wickham, 2016). The full reproducible code is available via the OSF repository.

1.6 Results

We removed a total of 7 participants from the analysis. Two participants were excluded for low accuracy (below 70%) in general comprehension questions and 5 participants were removed due to having over 18% trial “skip” rates, where their overall reading time of the stimulus was below 5 seconds (suggesting that they were simply pressing buttons as quickly as possible to avoid performing the reading task), this also removed a further 6 individual trials. This leads to a small imbalance in participant numbers, so we collected data from an additional two participants. With these elimination protocols 77 participants (96% of the data) remained for analysis. The mean reading times are presented in Table 2.

1.6.1 Reading time of Stimuli

Firstly, we investigated the stimulus reading time (mean for non-spatial SRT=38532*ms*, SD=40560.94*ms*; mean for spatial SRT =43233*ms*, SD=96461.76*ms*) are presented in Table 2. Further, we investigated SRT and its relation to the stimulus place within the Block, and Block Type (BT: Unambiguous First vs Ambiguous First) , Subject Type (ST: Spatial vs non-Spatial) and Question Type (QT: Ambiguous vs Unambiguous). The LMM results are shown in Table 3. This indicates no significant difference in stimulus reading time between spatial and non-spatial texts. However, SRT for Ambiguous questions was significantly shorter if the first Block was Unambiguous. We also see a generally shorter SRT when the first block of questions was unambiguous, as well as shorter SRT for Ambiguous questions overall. Further, we found significantly higher SRT for ambiguous questions when ambiguous questions were asked first compared to when unambiguous questions were asked first, though this is also the main interaction for block type. Lastly, we found significant differences between spatial and non-spatial stimulus types but only when accounting for BT and QT, this interaction is clearly shown in Figure 3.

Table 2.

LMM results for SRT as a factor of ST,BT and QT the random factors included were item,subject and age.

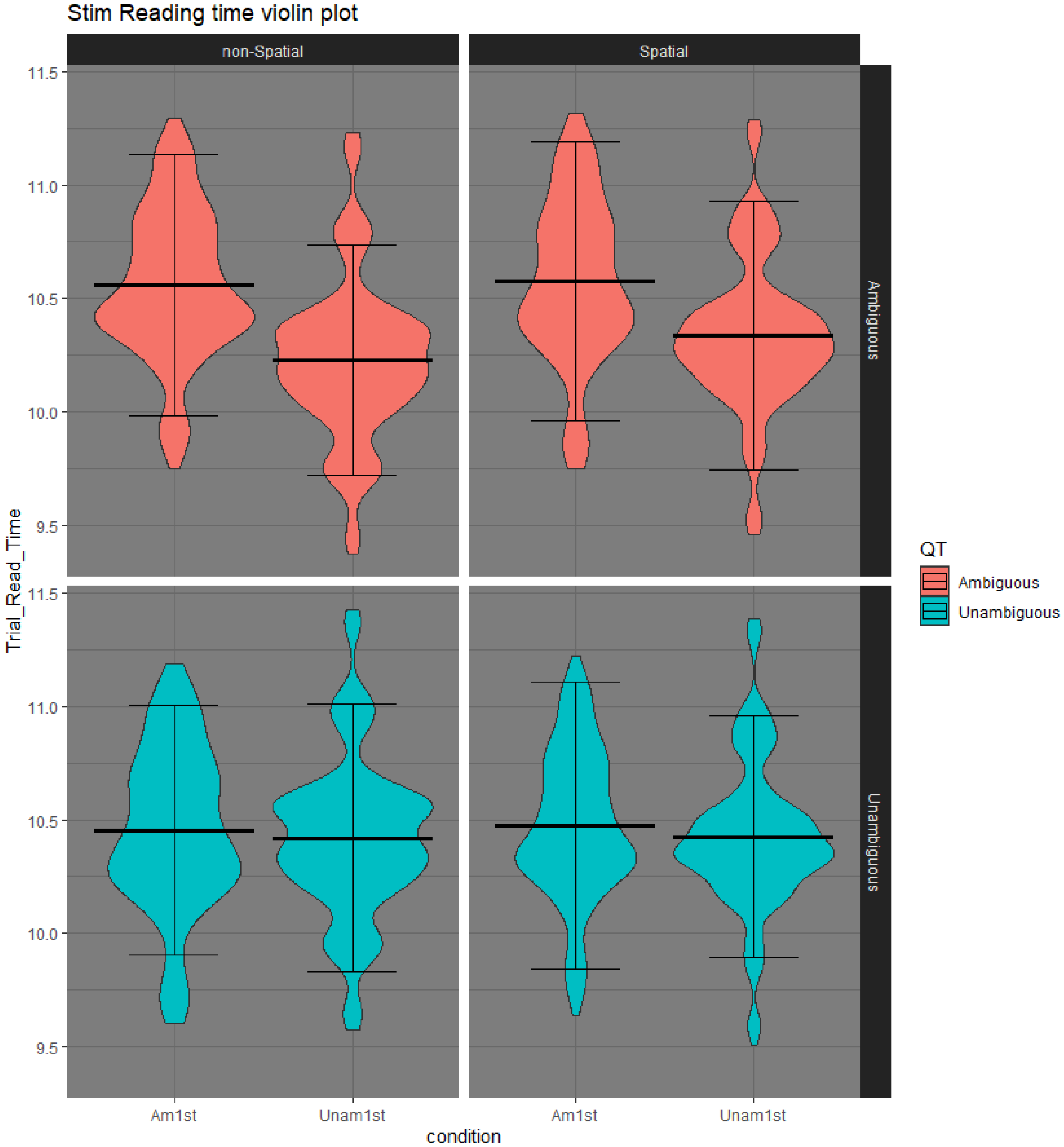
Fixed effects	**Stimulus Reading Time (log(RT) in ms**				
	b	**SE**	**df**	**t value**	**Pr(>|t|)**
Intercept	10.570	0.053	103.30	199.45	**<.001**
ST	-0.011	0.031	370.17	-0.38	.704
BT	-0.316	0.083	102.87	-3.809	**<.001**
QT	-0.136	0.029	3924.47	-4.627	**<.001**
ST*BT	0.073	0.046	3926.22	1.577	.115
ST*QT	0.056	0.042	3940.11	1.316	.188

BT*QT	0.325	0.044	3912.41	7.294	**<.001**
ST*BT*QT	-0.165	0.063	3915.64	-2.611	**.0091**

*<u>Note:</u> Final converged model ST*BT*QT+(1|item)+(1|subject)*

<u>Figure 3.</u>

Predictive GLMM model plot for stimulus reading time, showing interaction between ST,BT and QT.



1.6.2 Modulation of SRT by trial order

A point of interest was how trial number (Sequence (Seq)) within the blocks (Unambiguous/Anchor first vs Ambiguous first) could influence participants' reading behaviour in terms of SRT. We therefore investigated SRT as a function of trial order (Seq), ST and QT. If the reading of Situation models is non-deterministic we should find no modulation changes influenced by QT. ST, was included in the model in order to better aid in explaining deviance in stimulus reading time. We ran two models, one for each block type. We fitted a GAMM model investigating the differences between the two block conditions. The results show that the smooth term of sequence within both blocks was significant (edf= 1.147, F=7.46, p= .006 & edf= 1.00, F=5.15, p= .023) respective to each block, this is visualised in Figure 3. However, while the SRT decreases significantly along the sequence it appears that when the block type changes there is a significant increase in SRT. This increase then begins to decrease once again. We investigated the block changeover period (Transition) for both block orders by selecting only the 5 last trials of the first block and the first 5 of the second block and running an LMM. Due to the general downward trend of SRT across trials we used 5 trials before and after the changeover in order to investigate influence on SRT. The LMM investigates SRT as a function of BT and Block-Number (First Block 0, Second Block 1). The results (shown in Table 4) indicate that there is a significant difference in SRT between Blocks and Block Type. When the Ambiguous questions are asked in block 1 the block transition causes SRT to be significantly higher in block 2 when compared to when the Unambiguous Questions are asked first. Figure 5 shows this interaction in the form of a violin plot. The smoothing parameters, LMMs and full reproducible code can be found on the OSF in the “Scripts” folder.

Figure 4.

Predictive GAMM model showing Trial reading by (smoothed) BT&QT over trials. The thick black line indicates the block change.

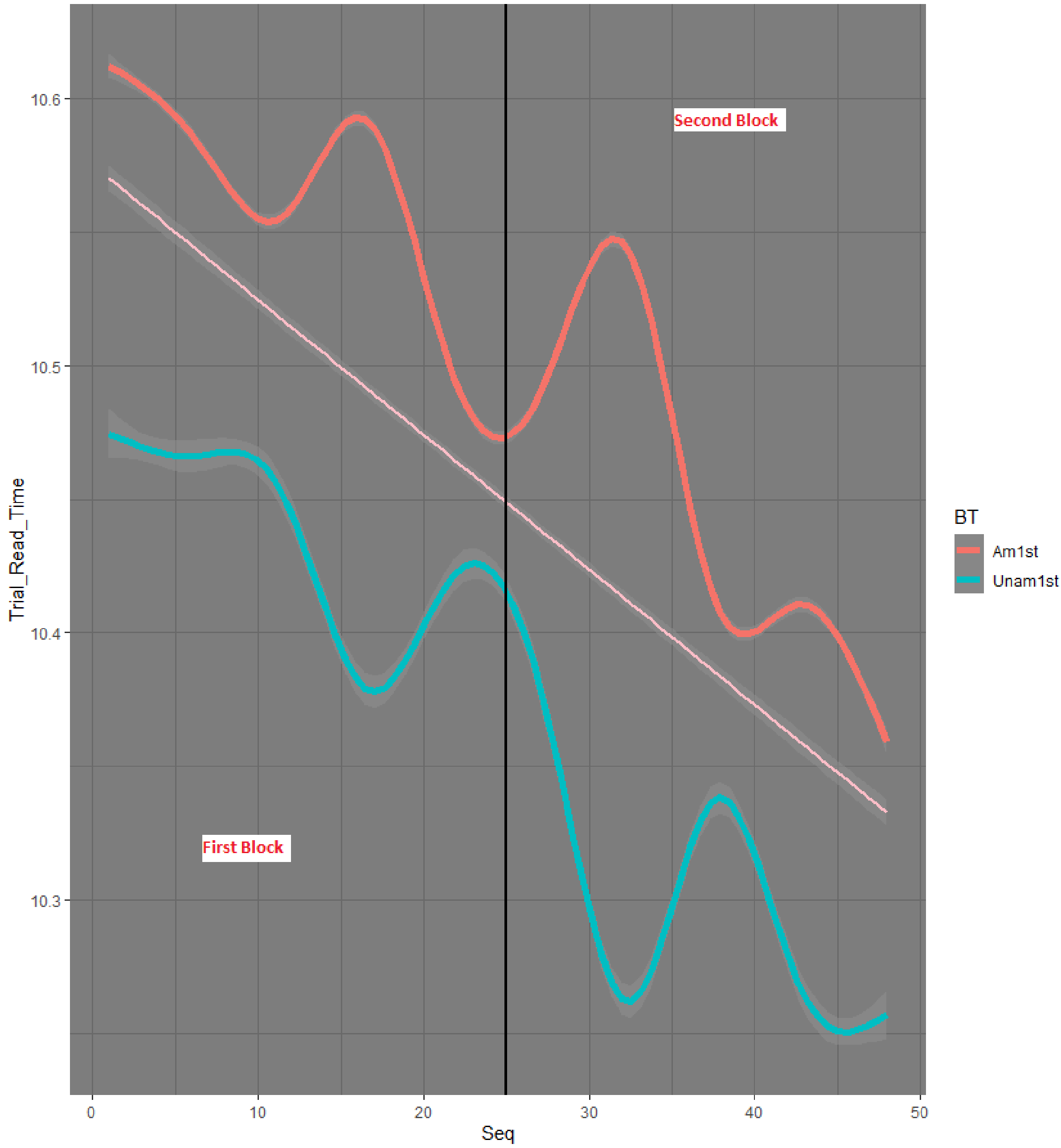


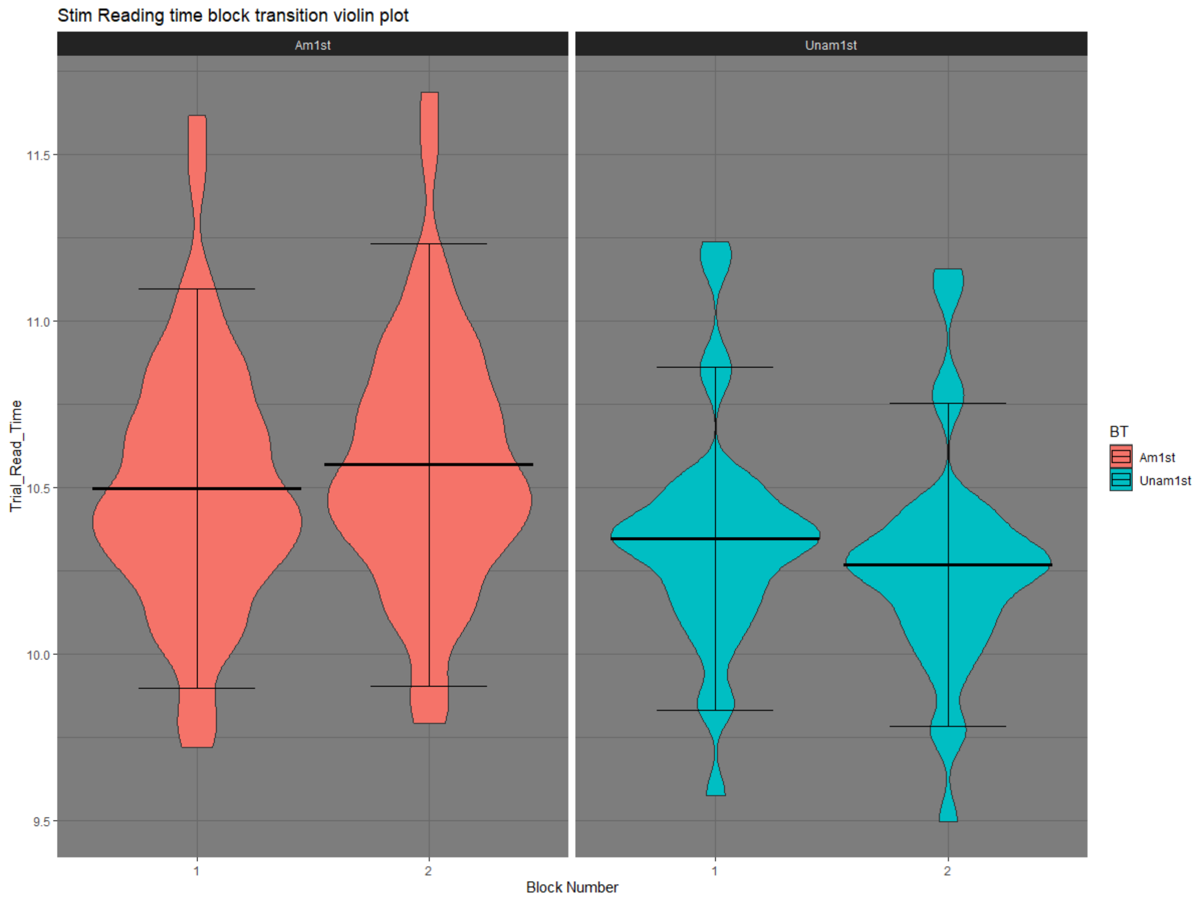
Table 4.

LMM results for SRT as a factor of ST,BT and QT. Including only trials from 20-30, the random factors included were item,subject and age.

Fixed effects	**Stimulus Reading Time (log(RT) in ms**				
	b	**SE**	**df**	**t value**	**Pr(>|t|)**
Intercept	10.496	0.059	96.840	175.22	**<.001**
Second Block	0.069	0.037	832.29	1.82	.068
BT	-0.150	0.094	97.14	-1.59	.113
Second Block*BT	-0.146	0.059	833.46	-2.47	**.013**

Figure 5:

Block transition change in trial reading time for Ambiguous questions first (Red) and Unambiguous first (Blue), Block number 1 refers to 5 trials prior to block transition and Block number 2 refers to 5 trials after block transition.



1.6.3 Question Reading time

We investigated Question reading time (QRT) as a function of ST,BT and QT. The QRT means and standard deviations are shown in Table 5. We find several significant differences in QRT. Question Type led to significantly lower QRT while Subject Type led to significantly higher QRT, indicating that participants are quicker to answer unambiguous questions while questions probing spatial models take longer for participants to answer compared to non-spatial ones. Further, we find a significant interaction between Block Type and Question Type whereby, if the first block contained ambiguous questions, it had little effect on the QRT of participants when entering the second (Unambiguous) question block, however when starting with unambiguous questions first participants had significantly longer QRT when entering the ambiguous block. This is further explored in the “Modulation of QRT by trial order”. Further, we find that ambiguous spatial questions took significantly longer than ambiguous non-spatial ones. The results of the LMM can be seen in Table 6.

Table 5.

Means Table of Question reading time

Mean QRT of conditions			
Ambiguous First	QT	Stimulus Type	log(QRT)
	Ambiguous	non-Spatial	9.01(0.092)
	Ambiguous	Spatial	9.05(0.087)
	Unambiguous	non-Spatial	8.98(0.070)
	Unambiguous	Spatial	9.03(0.074)
Unambiguous First			
	Ambiguous	non-Spatial	9.01(0.073)
	Ambiguous	Spatial	9.03(0.072)
	Unambiguous	non-Spatial	9.00(0.063)
	Unambiguous	Spatial	9.05(0.057)

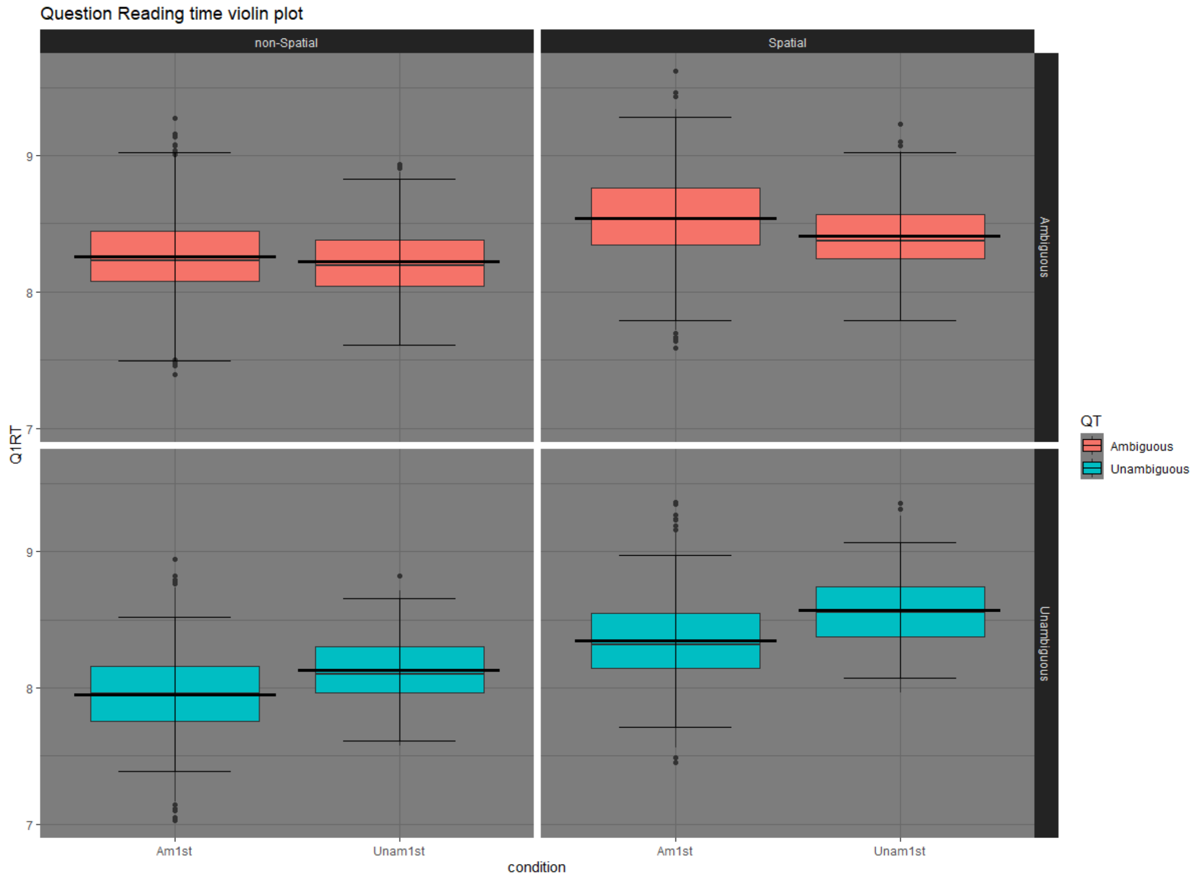
Table 6.

LMM results Table showing Log (Question reading time) as a factor of BT,QT and ST

Fixed effects	**Question Reading Time (log(RT) in ms**				
	b	**SE**	**df**	**t value**	**Pr(>|t|)**
Intercept	8.28	0.052	139	157.76	**<.001**
BT	-0.067	0.075	116	-0.90	0.369
QT	-0.363	0.034	3940	-10.40	**<.001**
ST	0.229	0.047	125	4.84	**<.001**
BT*QT	0.294	0.052	3920	5.59	**<.001**
BT*ST	-0.0237	0.052	3910	-0.45	0.651
QT*ST	0.211	0.051	3960	4.10	**<.001**
BT*QT*ST	0.0014	0.076	3940	0.019	0.985

Figure 6.

Predictive plot based on LMM results shown in Table 6, Log (Question reading time) as a factor of BT,QT and ST

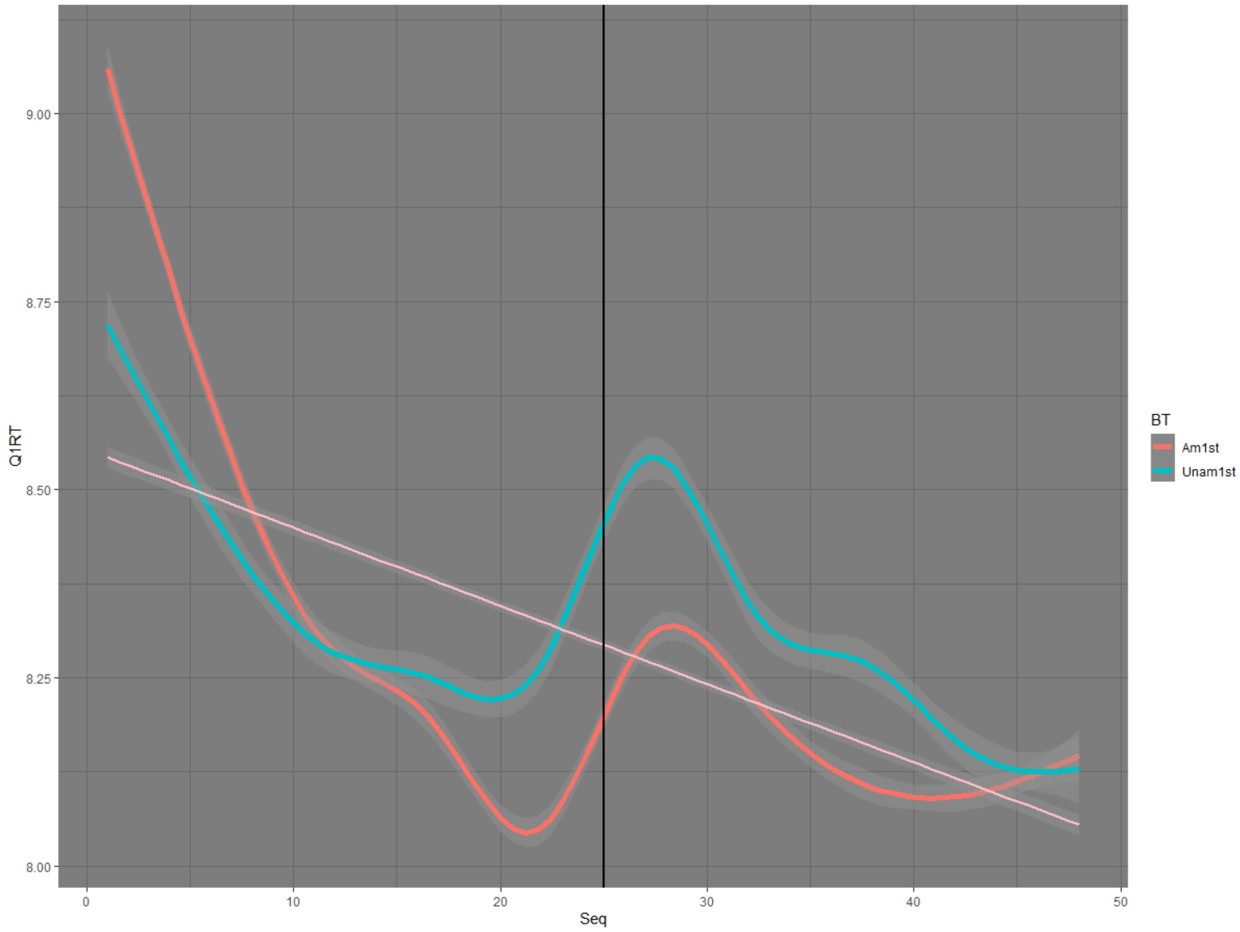


1.6.4 Modulation of QRT by trial order

We repeated the process of investigating trial modulation of SRT over trials in terms of QRT. We again ran a GAMM model with the same smoothing parameters as shown in 1.6.2. The model again included ST in order to better explain deviance. Our aim with this was to find out if the same “learning” reaction occurred in both SRT and QRT. Using Sequence within the blocks (Unambiguous/Anchor first vs Ambiguous first) and QT we determined that the transition between blocks caused significant differences in participants' question reading time. This is shown clearly in Figure 4.

The results show that the smooth term of sequence within both blocks was significant (edf= 9.470, F=13.36, p= <.001) , owing to the downward trend of QRT across trials visible in Figure 4. We also again find that the smoothing parameter of BT is significant (edf= 11.51, F=4.93, p= <.001). QT is not significantly different (p= 0.15) which may indicate that the transition from one block into the other causes the same general reaction regardless of question type. In order to better show this transition, we again ran an LMM for only the 5 trials before and after the transition. The results of this are shown in Table 6. The LMM investigates QRT as a function of BT and Block-Number (First Block 0, Second Block 1). The results indicate that there is a significant difference in QRT between Blocks and Block Type. In both Block types, QRT is significantly higher in the first 5 trials of the second Block compared to the last 5 of the first . Despite an overall downward trend in QRT across trials The smoothing parameters, LMMs and full reproducible code can be found on the OSF in the “Scripts” folder.

Figure 7.



Predictive GAMM model showing Question reading time by (smoothed) BT & QT over trials.

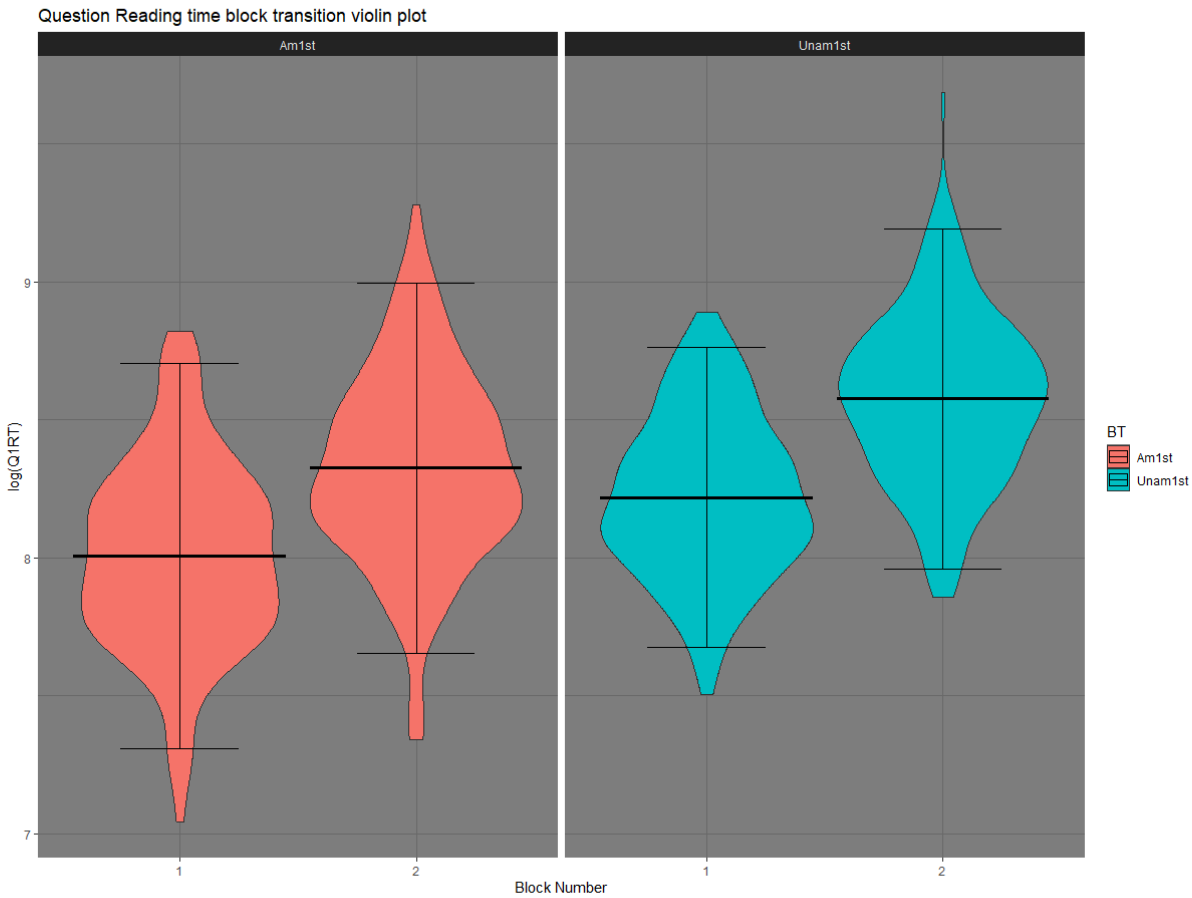
Table 6.

LMM results Table showing question reading time of trials from 20-30 as a function of BT and Block number

Fixed effects	**Question reading time (log(RT) in ms**				
	b	**SE**	**df**	**t value**	**Pr(>|t|)**
Intercept	8.03	0.063	149	126.992	**<.001**
Second Block	0.270	0.046	815	5.84	**<.001**
BT	0.184	0.088	128	2.092	**<.001**
Second Block*BT	0.0821	0.072	812	1.134	0.257

Figure 8.

Block transition change in question reading time for Ambiguous questions first (Red) and Unambiguous first (Blue), Block number 1 refers to 5 trials prior to block transition and Block number 2 refers to 5 trials after block transition.



1.6.5. Question Accuracy

Furthermore, we investigated question accuracy. Table 7 shows descriptive statistics for question accuracy . While it is possible to see the effects of ST and ambiguity at a glance from the means Table we ran a GLMM, using ST, QT and BT as predictors for question accuracy. Table 8 shows the results of this GLMM. It is possible to see several significant interactions influencing question accuracy. Firstly, we find that question type significantly influenced question accuracy whereby unambiguous questions had significantly higher accuracies than ambiguous ones. Further, the results indicate that questions probing spatial models had significantly lower accuracy than non-spatial probes, and while unambiguous questions had significantly higher accuracy overall, this effect was significantly more pronounced in non-spatial unambiguous questions. We also find a significant interaction between block type and question type, whereby the block transitions from Ambiguous to Unambiguous cause a significant increase in question accuracy post transition while a transition from Unambiguous to Ambiguous causes a significant decrease.

We further explore the effects of the block type transition in section 1.6.6 using GAMM models.

Table 7
Means Table of Question accuracy

Mean QRT of conditions

Ambiguous First	QT	Stimulus Type	Accuracy %
	Ambiguous	non-Spatial	0.496(0.500)
	Ambiguous	Spatial	0.467(0.499)
	Unambiguous	non-Spatial	0.819(0.385)
	Unambiguous	Spatial	0.397(0.489)
Unambiguous First			
	Ambiguous	non-Spatial	0.473(0.499)
	Ambiguous	Spatial	0.398(0.490)
	Unambiguous	non-Spatial	0.816(0.388)
	Unambiguous	Spatial	0.505(0.500)

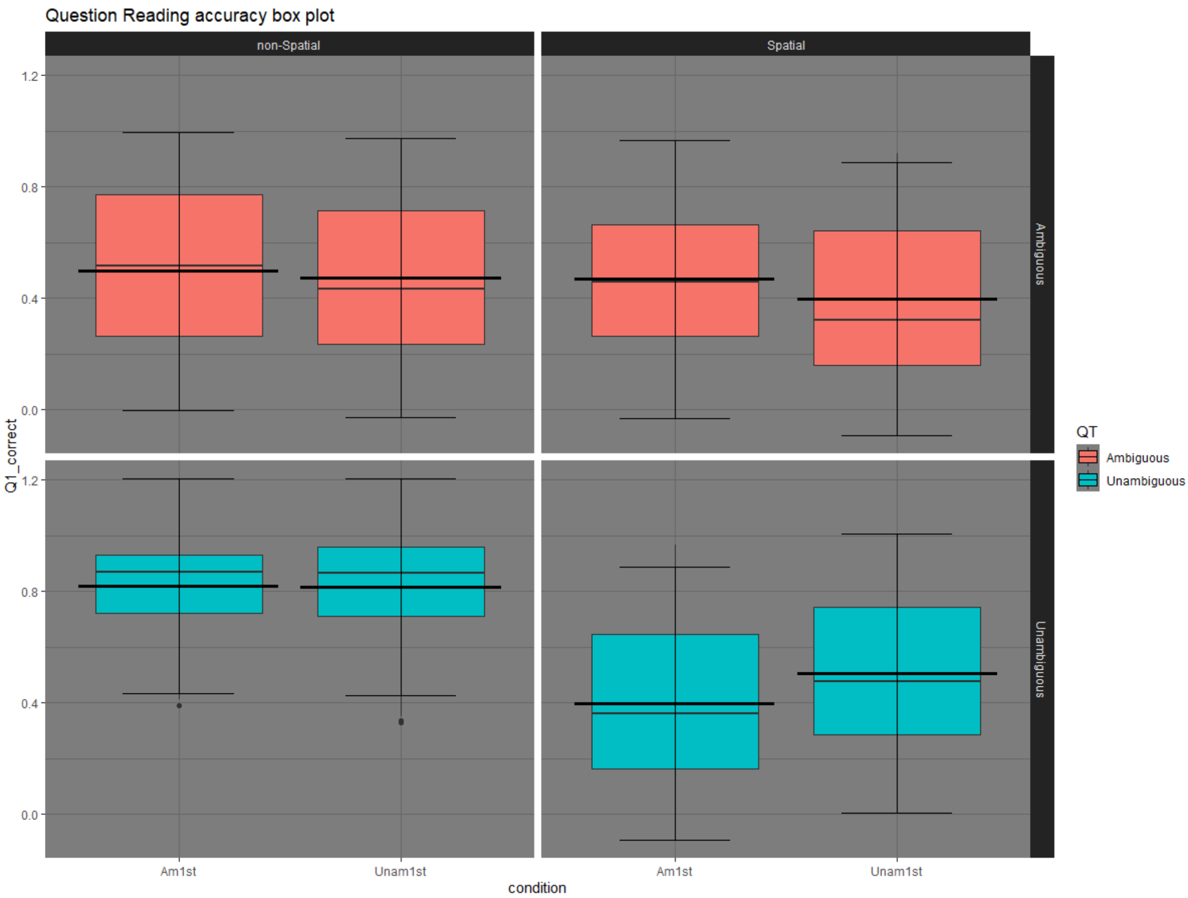
Table 8.

GLMM results Table showing question accuracy (%) as a function of BT,QT and ST

Fixed effects	**Question accuracy**				
	b	**SE**	**df**	**t value**	**Pr(>|t|)**
Intercept	0.219	0.261		0.84	.401
BT	-0.315	0.381		-0.828	0.408
QT	1.114	0.166		6.693	**<.001**
ST	-0.951	0.201		-4.72	**<.001**
BT*QT	0.991	0.255		3.884	**<.001**
BT*ST	0.455	0.225		2.025	**.043**
QT*ST	-0.787	0.226		-3.473	**<.001**
BT*QT*ST	-0.654	0.342		-1.911	.056

Figure 9.

Predictive boxplot showing the interaction between block type, question type and subject type.



1.6.6 Modulation of Question Accuracy by trial order

We again investigated the effects of trial sequence across the experiment in order to investigate participants “learning”behaviours” and to see how this was influenced by the block transition as well as QT. We ran a single GAMM model using Sequence, BT across sequence and QT across sequence as main smoothing predictors. We included ST across sequence, Subject (Participant) and Items as random smoothing terms in order to better explain deviance.

The results show that the smooth term of sequence within both blocks was significant (edf= 1.056, F=5.766, p= <.001) , Indicating a significant change in question accuracy from pre to post BT transition . We also found that the smoothing parameter of QT over sequence is significant (edf= 4.424, F=16.96, p= <.001) indicating that question types are significantly different from another across trials. The fitted GAMM models can be seen in Figure 6. We also wished to further investigate the difference in question accuracy pre and post block transition and so we performed the same data manipulation as found in 1.6.2 and 1.6.4. We used data only 5 trials prior and 5 trials post the block change and ran a GLMM in order to better understand the effects of the Block transition on question accuracy. The GLMM investigates QRT as a function of BT and Block-Number (First Block 0, Second Block 1). The GLMM indicates that, both the block type and block number significantly influenced question accuracy, whereby starting in the ambiguous block caused a significant increase in accuracy when moving to the second, unambiguous block, while the reverse of the (moving from unambiguous to ambiguous) causes a significant decrease in accuracy.

Figure 10.

Predictive GAMM model showing question accuracy (%) as a function of (smoothed) ST,BT,QT over trials.

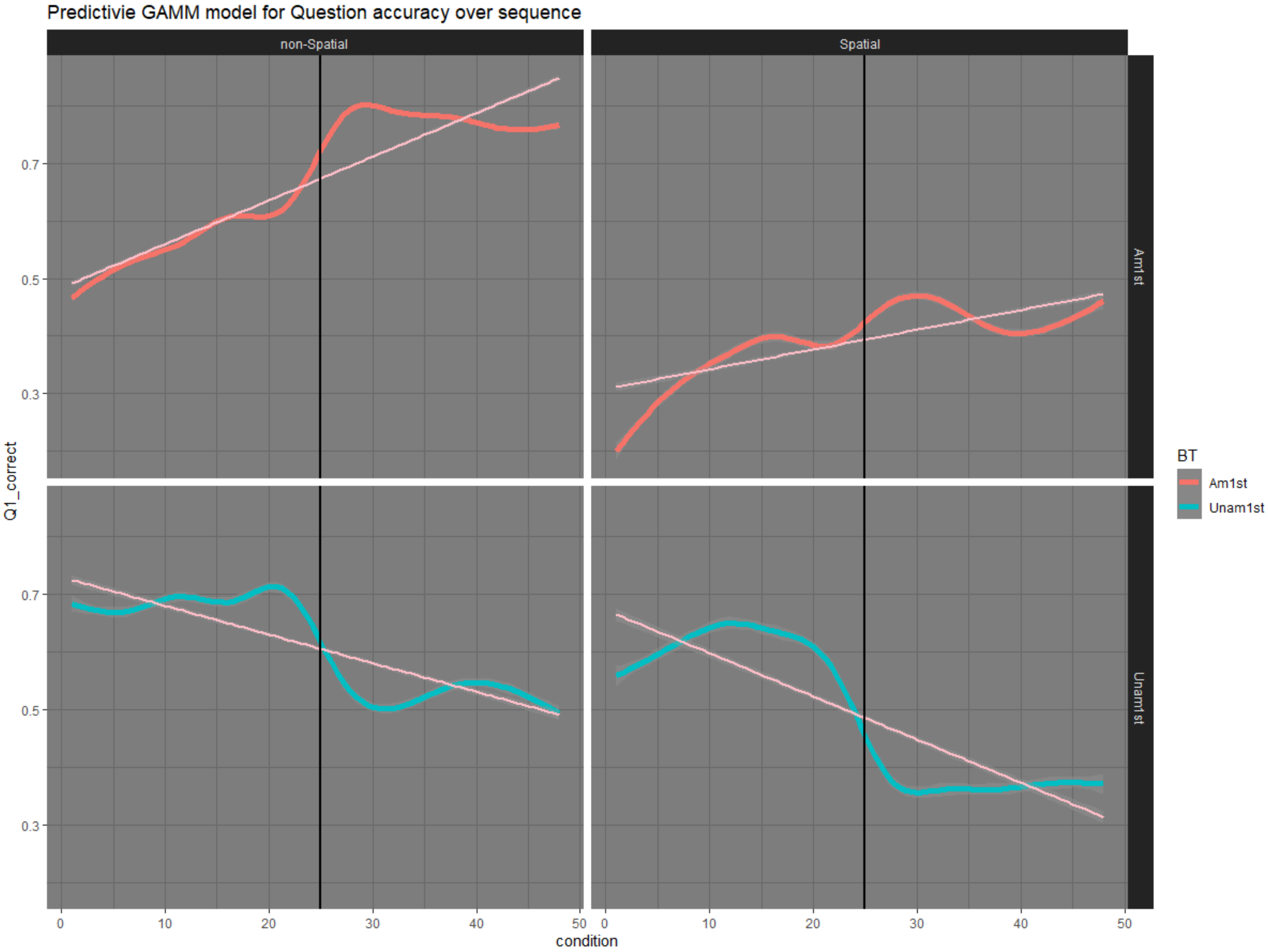


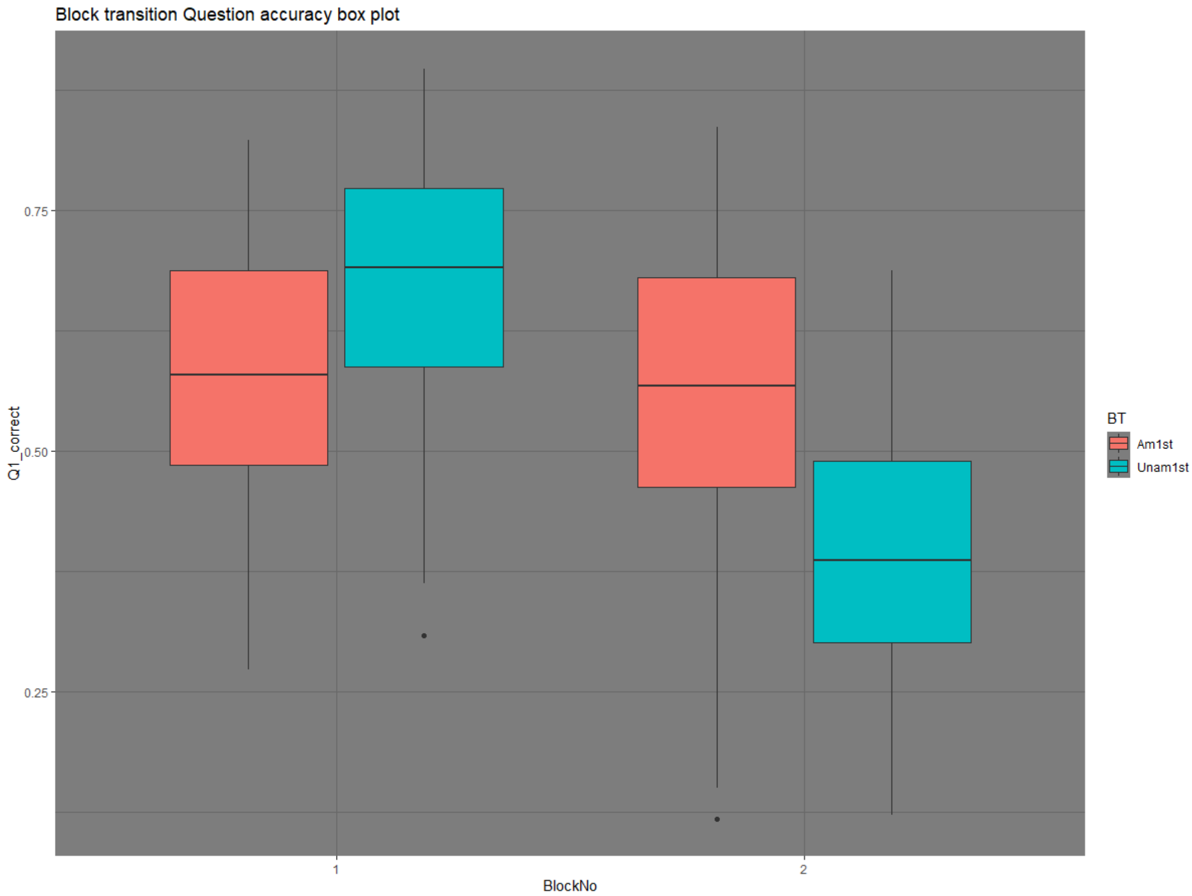
Table 9.

GLMM results for Question accuracy of trials from 20-30 as a function of Block number and BT.

Fixed effects	**Question reading time (log(RT) in ms**				
	b	**SE**	**df**	**t value**	**Pr(>|t|)**
Intercept	0.380	0.001	-	345.84	**<.001**
Second Block	0.061	0.001	-	56.11	**<.001**
BT	0.7420	0.001	-	674.43	**<.001**
Second Block* BT	-1.690	0.001	-	-1535.3	**<.001**

Figure 11.

Block transition change in question accuray for Ambiguous questions first (Red) and Unambiguous first (Blue), Block number 1 refers to 5 trials prior to block transition and Block number 2 refers to 5 trials after block transition.



1.6.7 Premise Line by line reading time

In order to better understand reading behaviour within a given trial, we investigated the time participants spent reading lines containing a premise. Specifically we wanted to know whether or not premise reading time differed between block types, and if reading time differed between premise 1 and premise 2. We ran an LMM based on line reading data for lines which contained a premise influenced by Premise (Prem 2) and Block type (BT). The results of this LMM are shown below in Table 10. Here we found no significant changes in reading time.

We also investigated whether the same factors (BT and Premise) influenced the re-reading likelihood of participants. To do this we ran a GLMM, the results of which are shown in Table 11. We found that Block type significantly changed re-reading rate whereby Participants encountering the Unambiguous question block first were significantly less likely to re-read the premises than those participants who encountered the Ambiguous question block first. Further, we found that in both Block types, Premise 2 was read significantly less often than premise 1.

Table 10.
LMM results for Premise line reading time in (log) ms as a factor of Block type and Premise

Fixed effects	**Premise line reading time (log(RT) in ms**				
	b	**SE**	**df**	**t value**	**Pr(>|t|)**
Intercept	7.46	0.111	5.860	67.33	**<.001**
BT	0.0244	0.477	91400	0.511	.609
Prem 2	0.00853	0.026	11900	0.328	.743
BT*Prem2	-0.0730	0.042	1190	-1.738	.082

Table 11.

GLMM results for Premise line re-reading likelihood as a factor of Block type and Premise.

Fixed effects	**Premise re-reading rate**				
	b	**SE**	**df**	**t value**	**Pr(>|t|)**
Intercept	-0.0539	0.250		-0.216	.829
BT	-0.406	0.097		-4.17	**<.001**
Prem 2	-0.112	0.049		-2.25	**.025**
BT*Prem 2	-0.0803	0.081		-0.979	.327

1.6.8 Premise Line by line re-reading and reading time by trial order

In order to better understand the line by line reading behaviour of participants, we investigated line reading time and Line re-reading rate by Block type and trial sequence (Seq) using GAMM models. We again added ST interacting with Seq as an additional smoothing parameter to better explain deviance. For consistency, models used the same smoothing parameters and knots as the previous GAMM models in this study. The smoothing parameters, models, visualisations and full reproducible code can be found on the OSF in the “Scripts” folder.

Firstly we found that for premise line reading time the smoothing parameter of Sequence and Block type was significant (*edf=.016, F=4.047, p=0.042*). This indicates that premise line reading time is different between the two block types, which is visualised in Figure 7.

Secondly, we found that premise re-reading likelihood was also influenced by Block type and trial sequence. The GAMM smoothing parameter indicated that this was significant (*edf=3.65 F=2.76 p=0.011*). However, the model also indicated that Block Number was not significant (p=.54) which indicates that the block transition did not significantly alter line re-reading likelihood. The line re-reading likelihood GAMM model is visualised in Figure 8.

Figure 7.

Predictive GAMM model of Premise line re-reading time by sequence

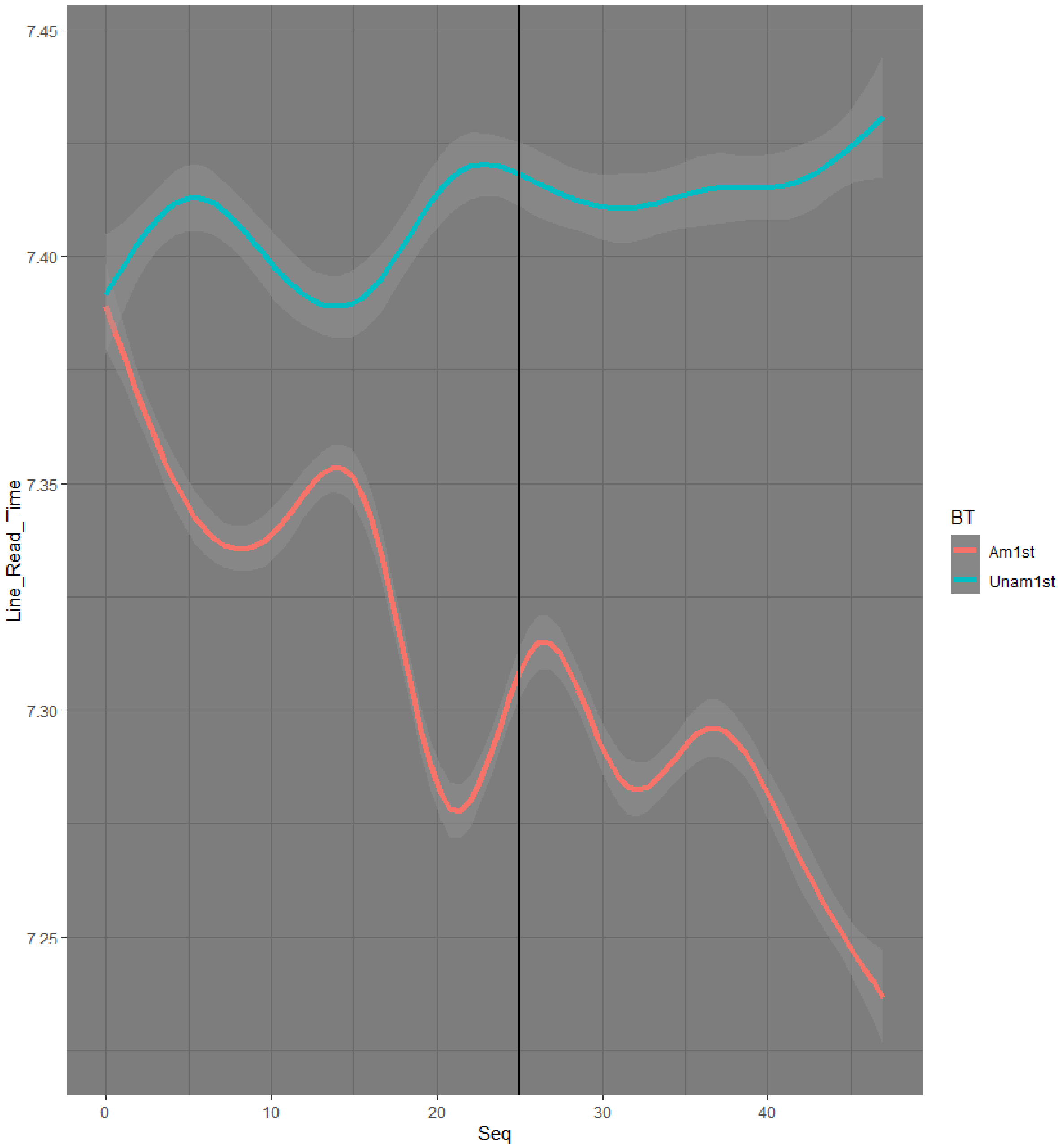
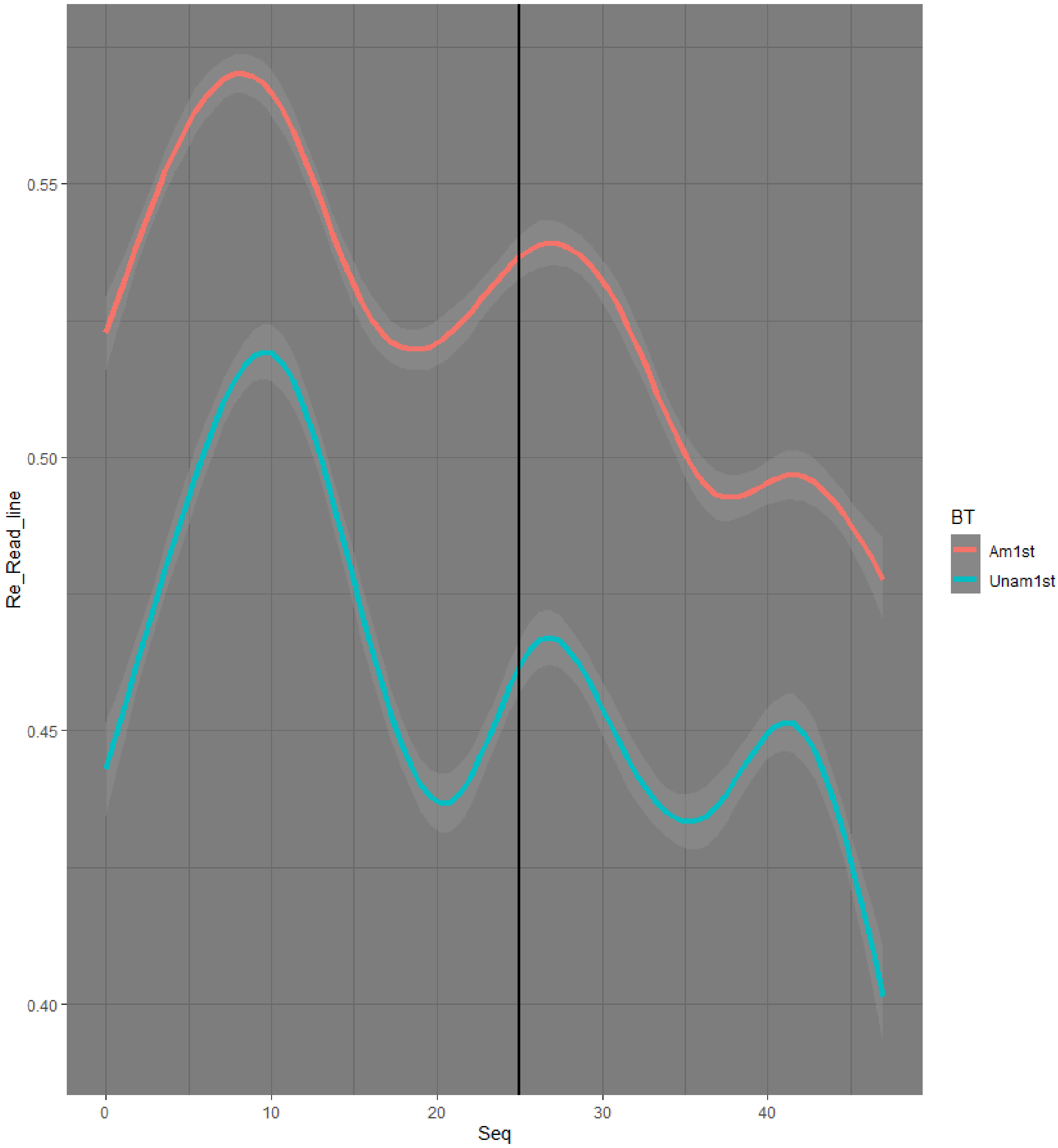


Figure 8.

Predictive GAMM model for premise line re-reading likelihood by sequence



Discussion

In this study we explored how ambiguous situational models are interpreted in terms of Stimulus reading time, Question reading time, Question accuracy and Line by Line Reading Time. We investigated how these were influenced by subject matter, the type of question asked, and a block manipulation of the question type. Further, we investigated how these variables changed over the course of the trial. The key results are as follows:

Firstly, in concordance with previous findings in Chapters 4 and 6, we did not find significant differences in SRT between spatial and non-spatial stimuli. However, all stimuli in this study described an ambiguous model, therefore it is important to note that the differences and non-differences may only reflect Stimulus reading times for ambiguous situational models. Furthermore, we found a significant difference in SRT based on the question type that followed it, as well as the Block Type. While it is possible that the stimulus reading time difference in BT is a result of the two different groups of participants having significantly faster/slower reading times from one another, this would not explain the differences seen as a result of the Question Type/Block Type interaction.

In section 1.6.2, we used GAMM models to investigate the modulation by trial order (Sequence), QT and BT. We found that there is a general increase in reading speed across conditions. However, when transitioning from the Ambiguous Questions to the Unambiguous Questions, participants significantly increased their SRT for a short time before trending downward again. We found a significant increase in SRT when transitioning from the last 5 trials of the Ambiguous Question Block to the first 5 of the Unambiguous question block. The fact that a change in SRT occurred moving from the ambiguous question block to the unambiguous one but not the other way around may indicate that, as we speculated, the holistic strategy causes readers to be more cautious in the construction and evaluation of their model to the extent that encoding is also influenced. The reverse however is not true, where the use of a preferred or “lazy” model, a generally less cautious approach does not change the way in which participants encode important information. Either a viable option created in a “lazy” way or simply a preferred mental model is useful to the participant in solving the task and are therefore considered to be “correct” (Nejasmic et al., 2015). As with all mental models the premises are “stored” in working memory (Goodwin & Johnson-Laird, 2005) but as they are considered to be the only viable option, are not revised (Rauh et al., 2005), which we have previously stated may be pivotal in ambiguity detection. Similarly to methodology in mental model research, participants may have created a singular viable option in which all premises held true (Jahn et al., 2007) but this preferred mental model (Franz & Garnham, 1990; Baddeley, 1998; Johnson-Laird & Byrne, 1991; Rauh et al., 2005; Goodwin & Johnson-Laird, 2005; Knauff & May, 2006) is not useful to detect ambiguity but remains useful for making other inferences.

Alternatively, the finding is consistent with reading time changes in task switching (Wallot,et al., 2019), in inference tasks. However, this is not likely, as the SRT change does not seem to occur when transitioning from the anchor probe block to the internal probe block. It is instead more likely that this “probe switching” effect comes from strategies which are conducive to ambiguity detection being generally more useful for both ambiguity detection and anchor locating. Something similar occurs in lexical/syntactic ambiguity detection (Swets et al., 2008), where participants without a need/strategy to detect ambiguity do not detect ambiguity at all. However, this assumes that lexical/syntactic ambiguity comprehension is similar in nature to situational ambiguity detection/comprehension, which may not be the case.

It would stand to reason that participants encouraged to be more cautious and create a more holistic interpretation of the situation in the first block become increasingly efficient at combining the two model premises and revising the model. The more efficient a participant is able to do this the more efficient they would become at detecting ambiguity, however once the block switches, participants are required to solve the unambiguous/anchor question. It is possible that this task switch encurs similar penalties as have been found in other task switching experiments (Masson et al., 2003; Wallot et al., 2019). This changes the focus to the unambiguous anchor (Point ***A***) rather than the ambiguous point (***B*** or ***C***), which are only described in relation to Point ***A***. A likely reason for the increase in SRT is that the participants have adopted a strategy for ambiguity detection during the first 24 trials, when the block switches they are forced to adjust their encoding to be able to locate the anchor of the model. Further, the ambiguous questions probe the “centre” of the model i.e “Find B, in the model ABC”. As such participants are unable to use only a single premise from which to infer B’s location. This discourages and penalises participants for using “lazy” or singular preferred models when reasoning situational models. This is consistent with our findings in Experiment 2 of Chapter 5.

While creating a “lazy” or singular preferred model might save time and can be used finding the anchor’s location in the unambiguous block, it is detrimental to ambiguity detection. As we will discuss shortly, ambiguity is not detected as efficiently in the Unambiguous first block type as in the other ambiguous first block type. It stands to reason that participants would be less inclined to adjust their reading behaviour (time spent reading the stimulus) to accurately memorise the information for the new task. In situational models it may therefore be the case that “teaching” participants with a specific task in a block may greatly influence their ability to perform another, for better or for worse. This kind of effect has been noted in prior research (e.g., Drieghe & Brysbaert, 2002; Wagner & Koutstaal, 2002), though it should be noted that this research used single word priming for word identification and not situation models. However, research into task switching and/or priming (Gilbert & Shallice, 2002; Papadpoulou & Clahsen, 2006; Schendan, 2017) have found effects similar to the ones we describe here. Furthermore, it is known that from mental models that reasoners use a variety of strategies but are able to adapt these if needed (Bacon et al., 2003).

We also found similar effects related to the block transition in question reading time. We found a significant influence on QRT by Question Type and Block Type. We found that unambiguous questions had significantly faster reading speeds than ambiguous ones. BT also significantly influenced QRT but only in relation to question type. Through the GAMM models presented in section 1.6.4 we can further interpret this finding. The model clarifies that in both block types, QRT significantly increases post block transition, but that this increase in QRT influences both block types in a similar way, something that is shown more clearly in the LMM presented in Table 6. This indicates that transitioning from questions probing the ambiguous part of the model to the anchor, and questions probing the anchor to the ambiguous part of the model are a likely cause for the increase in QRT. Participants, having learned to answer a specific question type, indicated by the significant downward trend across trials, for the past 24 trials, are faced with a question probing a different aspect of the model and must therefore adjust.

We also replicate previous findings from Chapters 4 and 6 that questions about spatial models take longer to answer than non-spatial ones. However, we find no significant differences in how spatial and non-spatial situations are handled in terms of whether the internal or anchor was probed. Therefore it is likely that our prior suggestion that semantic interference (Knauff & Johnson-Laird 2002;Knauff & May, 2006), causes additional difficulty when processing spatial situations and that this interference is dependent on what part of the model is probed. Instead we suggest that the difficulty caused by semantics affects the encoding (longer reading time) and processing (longer question reading/reasoning time) which is consistent with prior findings on semantic interference and mental models (e.g., Castelain & van der Henst, 2021; Lopiccolo & Chang, 2021). However, the precise mechanism and cause for semantic interference, and why this is more prevalent in spatial situations is unknown.

Lastly, we investigated question accuracy, in terms of BT, QT, ST and across trials. Concordant with prior findings we see that spatial stimuli had significantly lower question accuracy than non-spatial stimuli and we find that probing the ambiguous part of the model leads to lower comprehension than probing the unambiguous part of the model. This supports the reasoning that ambiguity detection is dependent on which part of the model is probed. Additionally, it is likely that the anchors of ambiguous models can be identified just as easily as those from unambiguous models. However, this points out a glaring flaw in our experimental methodology as we never probe the internal part or anchor of unambiguous models. This is something that should be rectified in future in order to ratify our findings on anchor location and ambiguity detection.

The main finding of the question accuracy analysis can be found in section 1.6.6, where we investigated the influence of trial order, and block types, on how accurately models are interpreted. Here we again find a significant change related to the block transition. Changing from Ambiguity probing questions to Anchor probing questions caused a significant change in accuracy and vice versa. In the case of transitioning from ambiguity probing to anchor probing we see a significant increase in question accuracy, for both spatial and non-spatial situations. While on its own this finding would support the notion that unambiguous models are easier to interpret than ambiguous ones, likely as a result of ambiguity only being detected upon the successful integration and evaluation of two premises.

We found, moving from ambiguity probing questions to anchor probing questions causes a brief increase in SRT and causes an increase in question accuracy, while moving from anchor to ambiguity does not cause an increase in SRT, and decreases accuracy. This supports our assertion that probing the internal part of the model encourages participants to adopt a more cautious and holistic approach leading to higher rates of ambiguity detection. An alternative explanation for this is that priming participants with ambiguous questions encourages them to not rely on “lazy” models or singular preferred models. In either case, it is known that reasoners of mental models construct models using a variety of strategies which depend on the task at hand (Schaeken et al., 2000), therefore it is likely that readers of situational models develop a strategy based on the question type. In our experiment a participant would have 24 trials to develop an efficient and accurate model encoding strategy. As well as an efficient and accurate reasoning strategy. Evidence for this is shown the generalised decrease in SRT over trials, generalised decrease in QRT over trials and increase in Block 1 question accuracy over trials. In the Ambiguity probing first block, these strategies may be specialised for ambiguity detection. Both strategies are likely to involve a clear understanding of both premises, i.e "***A*** in relation to ***B***" and "***A*** in relation to ***C***", allowing the participant to realise that ***B***'s relation to ***C*** is underspecified. This strategy would also work for locating the anchor, i.e object/subject "***A***" but with the drawback that the strategy is not time efficient when attempting to locate "***A***". As such a participant who began answering ambiguous questions might require additional reading time upon reaching the questions probing the anchor, in order to establish a more time efficient strategy, despite the prior strategy providing accurate anchor location.

Conversely a participant starting in the unambiguous block is likely to develop an accurate and efficient strategy for anchor object/subject location. Evidence for this is shown the generalised decrease in SRT over trials, generalised decrease in QRT over trials and increase in Block 1 question accuracy over trials. This strategy appears to be more time efficient than the strategy employed by readers in the ambiguous first block, as SRT is significantly lower, while also providing accurate results. A likely strategy employed here is the use of " good enough" processing, and using a single premise in order to accurately glean the location of the anchor point. Once a participant realises that "***A*** right of ***B***" and "***C*** left of ***A***", the participant uses the initial premise to determine that ***A*** is right of ***B***, since ***C*** is left of ***A***, ***A*** must be the furthest object to the right. It is only a matter of time before a participant realises that the location of the anchor ("***A***") can be determined easily by using only one of the two premises. In the case of the example, the first premise. This encoding strategy is highly time efficient as it keeps reasoning time to a minimum. However, the major drawback of such a strategy is that a reader is discouraged from using both premises, and as such does not realise that the relationship between ***B*** and ***C*** is underspecified. As such a transition from unambiguous/anchor questions to ambiguous/internal ones would be less likely to cause a participant to change their encoding strategy (SRT), though may make question reasoning more difficult temporarily. However, temporary the difficulty in reasoning (QRT Block Transition effect), if the model is not encoded correctly they are far more likely to get questions wrong.

While the method of this experiment does not allow us to accurately determine which encoding or reasoning strategy was used by participants, we have evidence to suggest that participants used different strategies in the different block types. In terms of encoding strategy we can use section 1.6.8 and Figure 7 to establish that the premise reading behaviours of participants were significantly different from one another between the different block groups. We see that participants presented with unambiguous questions first did not seem to change their reading time of the premises significantly, while participants presented with ambiguous questions first decreased their time reading the premises over the course of the trial. Using the results from Table 11 we also see that re-reading likelihood is significantly different between the two block types, and re-reading occurred far more frequently in the ambiguous questions first block. This may indicate that participants "taught" to detect ambiguity were more likely to re-read the premise lines of the stimulus, thus ensuring a better understanding of the model, as can be seen from the higher overall question accuracy. On the other hand participants in the unambiguous questions first block exhibit no-decrease in premise line reading time but do show a decline in re-reading rate, similar to that found in the other block type.

It may be that readers of the ambiguous first block are required early on to re-analyse (Rauh et al., 2005) the situational model in order to detect ambiguity accurately. This re-analysis of the situational model becomes more efficient as the experiment progresses (see Table 3, 1.6.2 & 1.6.8), and leads these participants in a better position to accurately interpret both ambiguous and unambiguous questions about ambiguous models. The participants presented with the unambiguous questions first may not have been "taught" to use this re-analysis strategy. While re-analysis is useful for detecting ambiguity and model comprehension it is not a requirement for efficient completion of the latter. As we suggested earlier, a time saving encoding strategy is preferred, which allows participants to encode premise information and gain an understanding of the model in accordance with the task required. However, this encoding strategy does not appear to change upon task change, i.e when the “find unambiguous anchor” task becomes the “ find the ambiguity” task. Instead we only see a change in QRT post change , which may indicate a change in reasoning strategy rather than an encoding strategy.

Conclusion

In this study we find evidence to suggest that unambiguous questions about ambiguous models are still more easily answered by readers than ambiguous questions. Our findings indicate that ambiguity detection and situational model interpretation involve different reasoning and encoding strategies which are dependent on question probing. We make some preliminary explanations as to the nature of the strategies in these tasks, however as this aspect of the research is exploratory, it is pertinent for us to further investigate these preliminary findings at a later date. However, we find strong evidence for the use of different, task dependent strategies in situational models. Furthermore, we replicate previous findings that Spatial situational models are harder than non-spatial ones.

Chapter 7: General Discussion and Conclusion

Throughout the studies in this thesis we investigated several research questions which had either been understudied or remained unaddressed. Firstly, the ability of readers to detect when a described situation has more than one viable model without being prompted has never been directly studied. Secondly, we investigated some of the differences in the reasoning of spatial and non-spatial situations.Thirdly, we explored how ageing influences not only the ability to detect ambiguity but also reason with spatial or non-spatial mental models. We investigated the eye-movements during the reading of ambiguous, unambiguous, spatial and non-spatial situations and what they might tell us about encoding strategies used by participants. Lastly, we investigated if ambiguity detection could be improved by probing the “internal” part of a model.

In all empirical chapters we found that participants find it harder to detect ambiguity and make inferences from ambiguous situations, compared to making inferences about unambiguous situations. The likely reason for this is that participants generally preferred to create one model from which to make inferences which is consistent with previous research (e.g., Barkowsky et al., 2005; Knauff et al., 2004; Nejasmic et al., 2015; Rauh et al., 2005). Creating a single mental model from which to reason is a useful strategy (Nejasmic et al., 2015) so long as the goal is to make inferences from a model and not to detect ambiguity. We found that participants may have had difficulty in constructing a single model from an ambiguous situation. This indicates that participants generally preferred to construct a single model from which to make inferences. However, as indicated by our findings in Chapter 5, participants begin to detect ambiguity more frequently in the latter half of the trials. Along with our findings in Chapter 6, a participant’s ineptitude at detecting ambiguity is likely a result of constructing a single model from which to reason. This model is initially “useful” for the reader to answer our questions about the model. However, through model re-evaluation (Rauh et al., 2005), it becomes apparent to the reader that the model they have consutrcted is insufficient and additional models need to be constructed. Leading to a higher rate of ambiguity detection in the latter parts of the studies. Furthermore, a higher question reasoning time indicated potential revision difficulty (Rauh et al., 2005) if the situation was ambiguous. It is possible that revisions of a single preferred mental model (Rauh et al., 2005), are not conducive to detect ambiguity but rather create another, viable model. Because mental models are constructed "bit-by-bit" (Johnson-Laird & Byrne, 1991), where the first premise forms the basis of the model while the second premise is "inserted" into the first, any revision to the full model results in a viable model and so long as this model does not violate either premise. The construction of a preferred, be it revised or not, model is not useful for the detection of ambiguity making accurate detection significantly less likely. However, because of the way in which the model was probed, not asking directly if there was more than one interpretation but rather probing the external or internal part of their model, using a preferred mental model from which to make inferences was useful to answer (Nejasmic et al., 2015). In Chapter 6 we investigated this further by "training" the participants with "internal" probing questions vs. "training" participants with questions that probed the anchor. We found that "training" participants using the anchor probes resulted in significantly lower rates of ambiguity detection compared to training with an "internal probe". This indicates that the models participants constructed from the premises were heavily influenced by the task they were asked to perform. This "task dependency" in ambiguity detection is not new but has so far only been found in lexical processing studies (e.g.,Huang & Ferreira, 2021).

This finding has greater implications in the study of mental models and preferred mental models. For example, if participants create models that are "useful" for the task at hand but are then required to make different inferences about their models than what they have been "trained" to do, some findings from previous mental model reaseach may not be wholly accurate. By using questions that promote participants to construct models that are overall more representative of the described situation, is beneficial when investigating how mental models are constructed.

We also compared the differences between inferences made about spatial and non-spatial situations and how ambiguity detection might be different between them. Across this thesis we found strong evidence for spatial situations being harder to reason with than non-spatial ones. While it is known that spatial reasoning may involve different strategies from non-spatial reasoning (e.g., Ford, 1995; Bacon et al., 2008) such strategies have not been extensively studied in mental model reasoning tasks and not at all in terms of ambiguity detection. In Chapter 2 we found significant differences in the ability of participants to accurately make inferences about spatial situations and detect ambiguity in spatial situations. We suggested numerous reasons for the difficulty in spatial reasoning. We must dismiss the findings from Chapter 2 however due to the involved an additional “4th actor” which may have inadvertently increased the number of entities within the model, known to cause increased difficulty in reasoning (Goodwin & Johnson-Laird, 2005). However, we found that even when this hurdle of the 4th actor was removed, spatial situations still had lower rates of comprehension and ambiguity detection (Chapter 5 & 8). We also suggested that the comprehension difficulty when making inferences about spatial situations found in Chapter 2 may have come as a result of the high working memory load of participants having to remember two models, made of four premises.

It is known that spatial reasoning is tied to greater working memory capacity needed (Friedman & Miyake, 2000; Zwaan & Radvansky 1998; Lin & Matsumi, 2022) and so by having participants needing to remember and reason with with more we have caused reductions in their ability to reason spatially (Gyselinck et al., 2007). However, the significantly lower spatial comprehension and ambiguity detection remained in later chapters where participants only had to deal with a single situation. We also suggested early on (in Chapter 2) that participants making inferences about spatial situations may have been influenced by semantics (e.g., Knauff & Johnson-Laird 2002;Knauff & May, 2006) which led to the creation of, in the case of unambiguous situations, an incomplete/incorrect model and ambiguity going undetected. A mental model, being a “visual” representation of objects/subjects constructed from a description, only contains each object/subject once (Johnson-Laird, 1994;Ragni & Knauff, 2013). While semantic interference problems in the reasoning of mental models have been found before (e.g., irrelevant descriptives, Dewar et al. 2007; Craig and Dewar, 2018), models from spatial and non-spatial situations are not compared. However, semantic interference has been known to also influence non-spatial reasoning processes (Goodwin & Johnson-Laird, 2005; Djiwandono, 2006;Román et al., 2013;Andrews et al., 2013;Román et al., 2015). We therefore cannot say for certain why spatial situation comprehension and ambiguity detection is significantly lower than non-spatial situation comprehension and ambiguity detection. However, we speculate that spatial situations undergo semantic “checks” to see if models obey laws of the real world (e.g., Goodwin and Johnson-Laird, 2005) more so than non-spatial situations. Not only would these “checks” take longer, often causing higher reading and reasoning time, but also make it significantly more likely for a descriptively incorrect (not obeying the premises) but globally correct (obeying unknown world,personal or semantic rules) model to be created from which to make inferences.

In regards to ageing we investigated differences in model comprehension, encoding (reading) behaviour (Chapter 4 and 6) and explored if older and younger readers used different strategies for encoding and reasoning. We found that, consistent with prior research older people differed significantly in their reading behaviour and reading time (e.g., Rayner et al., 2006). Older people took significantly longer to read the stimulus text as well as taking longer to read/reason when the question had been probed. While a longer reading and reasoning time known to occur in older readers, due to general cognitive slowing (Kemper et al., 1993; Liu et al., 2017), irrelevant information inhibition (Hasher & Zacks, 1988; Hamm & Hasher, 1992), difficulty in memory retrieval (Gerard et al., 1991) we instead argued that the longer reading and reasoning time may be down to compensatory strategies (Piefke et al., 2012) such as breaking down the text into smaller “chunks” (Swets et al., 2007). These compensatory reasoning or encoding strategies allowed older participants to remain on par with younger readers in terms of ambiguity detection and unambiguous model comprehension, which is consistent with prior research (e.g.,Light, 1988; Radvansky et al.,1990). Further, it is known that older people may have difficulty with integrating premises into a model (Copeland & Radvansky, 2007), but that this difficulty does not hinder their ability to make inferences from a model once it has been created (Gilbert et al., 2004). We suggested that older readers are "aware" of certain cognitive deficits such as lower working memory capacity (e.g.,De Beni et al., 2004; Fiore et al., 2012 ; Park et al., 2002) and compensate for this through additional re-reading (Rayner et al., 2006; Inhoff, et.al., 2017) of premises in particular in order to better encode these (Inhoff, et.al., 2017).

While we did not directly investigate which strategies people used to encode or make inferences from their answer, it is known from prior research that, for example, people use "their own" strategies to do so (Schaeken et al 2000). Furthermore, it is thought that older readers are able to use a more holistic reasoning strategy (Dror et al., 2005) which may aid in compensating for lower working memory capacity. Future research should investigate the precise strategies used by older and younger people to see if these are different due to the working memory deficits in older people.

Lastly we investigated if different reasoning strategies were used by participants to make inferences and detect ambiguity. We have already spoken about how older participants may have used compensatory strategies (e.g., frequent re-reading, Rayner et al., 2006) but we also wanted to know if by probing a different part of a model we could encourage or discourage certain reasoning strategies. In Chapter 5 we investigated if probing the "internal" part of the model i.e., "what entity is in the middle", to see if we could discourage the use of strategies not conducive to ambiguity detection. We found that when we probed the internal part of the model that both older and younger readers were able to detect ambiguity significantly better than when the "external" part of the model was probed. We suggested that by probing the internal part of the model we discouraged participants from using a single premise, akin to "good enough" reasoning in discourse processing

(Ferreira et al., 2002), a process leading to what we referred to as a “lazy” model. We also suggested that the creation of a preferred mental model (e.g., Rips, 1994; Johnson-Laird, 2001; Rauh et al., 2005) from ambiguous premises, remained useful (Nejasmic et al., 2015) to make inferences from and so ambiguity would often go undetected. As one of the viable models was preferred rather than one deactivated in favour of another (Rauh et al., 2005), this activated model would not undergo additional revision. However, by probing the internal part of the model we argued that participants became more aware of how the premises “fit” into the model as a whole rather than being constructed “bit-by-bit” (Johnson-Laird, 2001). Subsequently, probing the middle may have encouraged model revision (Rauh et al., 2005) when participants were asked about their model and, no longer being able to re-read the premises, used both premises stored in memory from which to reason/detect ambiguity. It is important to note that when we probed the internal part of the model, older readers outperformed younger readers in spatial situation comprehension, which may give further credence to our assertion that older readers are reliant on different strategies. A more holistic strategy, used by older readers to begin with (Dror et al., 2005), may have been the result of the internal probe.

We also investigated how probing the middle may influence a participant's ability to understand the situation as a whole. In Chapter 6 we found that by “teaching” participants to detect ambiguity during the first half of the experiment their ability to locate the unambiguous “anchor” was significantly better in the second part of the experiment. While the study in Chapter 6 may not be the best, having a few confounds and problems, we do show that participants can be encouraged to use a specific, more beneficial strategy to create mental models. We also found evidence to suggest that encouraging a reasoning strategy also influenced encoding behaviours,such as re-reading (e.g.,Inhoff, et.al., 2017) in the second part of the experiment. Further evidence to support our findings show that esults from more linguistic fields of cognitive psychology have described strategies that resolve ambiguities, at least temporarily. If a “good enough interpretation is re-vaulated this may not result in the correct interpretation (Huang & Ferreira, 2021, Slattery et.al 2013).

Readers tend to stick to the initial interpretation even when subsequent text indicates this interpretation cannot be valid (Christianson, 2016). A re-analysis of the premises may lead the reader to believe that the actual order of objects/subjects is more complex than initially thought but this may also not result in the correct interpretation that “there is not enough information”. Even when readers revise initial misinterpretations and arrive at the a new interpretation, ambiguities may remain unresolved, or undetected (Fujita & Cunnings, 2021; Huang & Ferreira, 2021; Meng & Bader, 2020; Slattery, et al., 2013), although this is speculative. When reading described situations, it is known that people use a variety of strategies for constructing and subsequently deducting a mental model (Schaeken, et al 2000)**.** If a reader determines “good enough” interpretations are a sufficient strategy for their construction of a model based on a described situation (as re-analysis may take more time and may not yield results), then it is likely that ambiguity in those situations could go unnoticed. Such lingering effects of ambiguity resolution may impact a reader’s ongoing construction of the situation model of the text.

Additionally, recent findings (Blott, et al., 2021) show that readers of garden path sentences (globally ambiguous sentences with more than one viable interpretation) do not always engage with a re-evaluation process of the sentence and follow a “good enough” strategy for interpretation of the sentence (Blott, et al., 2021). Blott et al., 2021, investigated lexically ambiguous sentences which were disambiguated towards the end of the sentence e.g “The ball was crowded”, readers were then asked to discern whether or not the sentence “made sense”. They found that readers often did not engage in re-evaluation and that when re-evaluation did occur that this took additional time. Further, it was found that a sentence that did not require re-evaluation had significantly shorter processing times than the ones that did. In the relation to situational ambiguity, the shorter processing time for unambiguous situations has not been consistently shown and may appear only when working memory is under a heavy load as we found in Chapter 2. The question used “does this sentence make sense?” takes a more direct approach to how ambiguity is detected than the studies in this thesis. However, it is possible that re-processing of a situational stimulus (or at the least the mental model it creates) incurs the same penalties associated with re-analysis found by Blott, et al., 2021. While the task of ambiguity detection of a mental model and Blott et al’s question “Does the sentence make sense?” may be different for participants, both tasks require revaluation which may be similar to model revision (Rauh et al., 2005). In non-model ambiguity, even when readers revise initial misinterpretations and arrive at the correct interpretation, effects of temporary ambiguities may linger (Fujita & Cunnings, 2021; Huang & Ferreira, 2021; Meng & Bader, 2020; Slattery, et al., 2013). It may be that a revised (Rauh, et al.,2005) model is simply the other viable solution and so encouraging participants to consider both premises as well as both solutions is required for ambiguity to be detected accurately. The same is also likely to be the case for syntactic ambiguity which has also been found to be task dependent (Swets, et al., 2008). Assuming that more linguistic text interpretations hold true when dealing with mental models. A reader may realise that a preferred model is not “good enough” but only after the text is no longer available and is thus forced to use the only model available to them., leading to lower ambiguity detection and higher QRT. Conversely, if the reader does realise that their understanding of the model is insufficient while the text is still available,subsequent reanalysis during stimulus reading may not yield the correct interpretation (Huang & Ferreira, 2021) or the original interpretation may be “held on to” (Christianson et al., 2017). By applying these described reasoning behaviours from psycholingustic language interpretations to mental models it is possible to move beyond the current knowns that participants are “blind” to alternative models (Johnson-Laird, 1994).

For instance, it is known that reading behaviours are influenced by the text which is being processed (Rawson, et.al., 2000; Margolin & Snyder, 2017; Wong & Moss, 2021). In contrast to this “teaching” participants to locate an unambiguous anchor, the location of which can be determined using only a single premise, does not aid them when detecting ambiguity. This shows that the model that is created is less useful for the understanding of the situation as a whole. However, we do find evidence that over time, the strategies used are adjusted. We suggest that not only is the creation of a mental model dependent on what task it is required for (Ambiguity detection or inferences), but also that the strategies used to create a mental model can be guided through careful probing. We have also established that the use of naturalistic paragraphs in the study of mental models is viable and could be used to further study mental model prefference. Not only does this have implications beyond ambiguity detection but also may elucidate how when and why ambiguity is detected by older people compared to younger ones.

Future research

It has to be noted that the experiments in this thesis were not conducive to determining the precise strategy used, for example using a preferred model (having been built from two premises) vs a “lazy” model approach. It is understood that reasoners use their own preferred strategies for creating and reasoning with mental models (Goodwin & Johnson-Laird, 2005; Nejasmic, et.al.. 2015). However, it is possible to tell that by encouraging or discouraging the use of a particular strategy it was possible to improve model comprehension accuracy and ambiguity detection. Naturally, this leaves a large gap in knowledge as it is currently unknown what specific strategies, e.g breaking down text into smaller chunks (Swets et al., 2007), and these should be investigated in the future.

It is important to keep these findings in mind when thinking of the potential future research questions that stem from them. For example, one limitation of the present work is that no working memory capacity data was collected. We frequently reference working memory capacity as a key component of ageing research, e.g., (Bopp & Verhaeghen, 2007; Borella et al., 2008; De Beni et al., 2004; Fiore et al., 2012 ; Park et al., 2002). We assume that our older participants have lower memory capacities compared to younger readers, but did not collect any data on this. It may be the case that some of our findings, like the ability to detect ambiguity, task switch and interpreting situational models may be better predicted by working memory than by age. Further, working memory may also predict model comprehension far better within older and younger groups.

In the same line of reasoning, the participants that partook the studies presented in this thesis all appeared to be relatively active readers. We frequently used online participant pools, while financially incentivised it is likely that those participants already have an interest in science, and are likely to be relatively avid readers. Furthermore, we often tested students from Bournemouth University, who are likely to be well educated and interested in science by default. Lastly we used the participant pool of the ADRC for our online and eye-tracking studies, again these individuals were likely to be interested in science and well educated. With this in mind it is possible to conclude that our findings may not be true for a general population and instead are a more of a sweeping generalisation about the educated, keen readers. Therefore a good line of enquiry would be to investigate a population which has lower years spent in education and are perhaps less keen readers. This information, in combination with reliable measures of working memory capacity may give further insights into how situational models are processed and how encoding strategies differ. Finally, our investigation presented in chapter 8 needs to be repeated using a between group design with older readers. If task switching is modulated by age/ working memory capacity then it could be that working memory capacity directly influences a person's ability to switch between tasks of a similar nature which require different encoding strategies. Additional repeats should also be run using comparative "neutral" texts, which do not contain a premise of any kind. This investigation would allow us to investigate the "risky" reading strategy in older readers without modulation from task dependency.

Conclusion

In conclusion, in this thesis we present a novel paradigm for investigating situational models, from a larger, more “naturalistic” text, which better resembles “real life” encounters with mental models. We found that ambiguity in a situation is hard for readers to detect but suggest that encouraging, or discouraging reasoning strategies allows participants to create a holistic view of the premises which is more conducive to ambiguity detection. Furthermore, we outline that different encoding strategies are better suited to specific tasks .We also investigate ambiguity detection and making inferences from mental models in ageing. We found that, consistent with prior research (e.g.,Light, 1988; Radvansky et al.,1990) that despite slower reading (Myerson et al., 1990) and processing (Copeland & Radvansky, 2007), older participants are just as capable at creating mental models and detecting ambiguity. We suggest that older participants may use more holistic reasoning strategies as compensatory mechanisms for lower working memory capacity. Our findings lead us to conclude that situational ambiguity is harder to detect for both older and younger people, and that this is highly task dependent, as probing different parts of a model causes different comprehension and ambiguity detection rates. We also suggest several reasons as to why spatial and non-spatial situational models are harder and easier, respectively, to comprehend. But acknowledge that the differences between spatial and non-spatial situations require significantly more work in future in order to better make a distinction.

Our findings lead us to conclude that situational ambiguity is harder to detect than simply solving a presented model, but that this is highly task dependent. Furthermore, we explored the concept that the “Risky” reading strategy adopted by older readers is beneficial to model comprehension, and in terms of re-reading, a similar strategy is adopted by younger readers, while older readers modulate their reading strategy in order to better comprehend the situational model described in the text. We found that older and younger readers only differ in terms of their general reading strategy but not in their comprehension of the models described and we assert that this is down to encoding strategy being task dependent. Further, different encoding strategies are better suited to specific tasks. We outline future lines of inquiry that will delve deeper into the influence of task dependency on reading strategy and can shed light just what strategies are adopted by readers performing different tasks.

References

Aks, D. J., Zelinsky, G. J., & Sprott, J. C. (2002). Memory Across Eye-Movements: 1/f Dynamic in Visual Search. *Nonlinear Dynamics, Psychology, and Life Sciences*, *6*(1), 1–25. https://doi.org/10.1023/A:1012222601935

Andrews, B., Aisenberg, D., d'Avossa, G., & Sapir, A. (2013). Cross-cultural effects on the assumed light source direction: Evidence from English and Hebrew readers. *Journal of Vision*, *13*(13), 2–2. https://doi.org/10.1167/13.13.2

Bacon, A. M., Handley, S. J., Dennis, I., & Newstead, S. E. (2008). Reasoning strategies: the role of working memory and verbal-spatial ability. *European Journal of Cognitive Psychology*, *20*(6), 1065–1086. https://doi.org/10.1080/09541440701807559

Baddeley, A. (1998). Working memory. *Comptes Rendus de l'Académie Des Sciences - Series III - Sciences de La Vie*, *321*(2–3), 167–173. https://doi.org/10.1016/S0764-4469(97)89817-4

Barkowsky, T., Freksa, C., & Knauff, M. (2005). Special Issue: "Spatial and Visual Components in Mental Reasoning About Space" Editorial. *Spatial Cognition & Computation*, *5*(2–3), 115–117. https://doi.org/10.1080/13875868.2005.9683799

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

Benoit, K., Watanabe, K., Wang, H., Nulty, P., Obeng, A., Müller, S., & Matsuo, A. (2018). quanteda: An R package for the quantitative analysis of textual data. *Journal of Open Source Software*, *3*(30), 774. https://doi.org/10.21105/joss.00774

Bharuthram, S. (2018). Facilitating active reading through a self-questioning strategy: student and tutor experiences and reflections of the strategy use. *Journal for Language Teaching*, *51*(2), 85. https://doi.org/10.4314/jlt.v51i2.4

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where Is the Semantic System? A Critical Review and Meta-Analysis of 120 Functional Neuroimaging Studies. *Cerebral Cortex*, *19*(12), 2767–2796. https://doi.org/10.1093/cercor/bhp055

Blott, L. M., Rodd, J. M., Ferreira, F., & Warren, J. E. (2021). Recovery from misinterpretations during online sentence processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *47*(6), 968–997. https://doi.org/10.1037/xlm0000936

Bopp, K. L., & Verhaeghen, P. (2007). Age-Related Differences in Control Processes in Verbal and Visuospatial Working Memory: Storage, Transformation, Supervision, and Coordination. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *62*(5), P239–P246. https://doi.org/10.1093/geronb/62.5.P239

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

Boudreau, G., & Pigeau, R. (2001). The mental representation and processes of spatial deductive reasoning with diagrams and sentences. *International Journal of Psychology*, *36*(1), 42–52. https://doi.org/10.1080/00207590042000001

Brébion, G. (2003). Working Memory, Language Comprehension, and Aging: Four Experiments to Understand the Deficit. *Experimental Aging Research*, *29*(3), 269–301. https://doi.org/10.1080/03610730303725

Brysbaert, M., & Stevens, M. (2018). Power Analysis and Effect Size in Mixed Effects Models: A Tutorial. *Journal of Cognition*, *1*(1). https://doi.org/10.5334/joc.10

Chen, J., Hale, S., & Myerson, J. (2003). Effects of Domain, Retention Interval, and Information Load on Young and Older Adults' Visuospatial Working Memory. *Aging, Neuropsychology, and Cognition*, *10*(2), 122–133. https://doi.org/10.1076/anec.10.2.122.14461

Christianson, K. (2016). When language comprehension goes wrong for the right reasons: Good-enough, underspecified, or shallow language processing. *Quarterly Journal of Experimental Psychology*, *69*(5), 817–828. https://doi.org/10.1080/17470218.2015.1134603

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., 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

Clifton, C., & Staub, A. (2008). Parallelism and Competition in Syntactic Ambiguity Resolution. *Language and Linguistics Compass*, *2*(2), 234–250. https://doi.org/10.1111/j.1749-818X.2008.00055.x

Crain. S, & Steedman. M. (1985). *On not being led up the garden path: the use of context by the psychological syntax processor* (D. R. Dowty, L. Karttunen, & A. M. Zwicky, Eds.). Cambridge University Press.

Craik, K. J. W. (1943). The nature of explanation. University Press, Macmillan.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*(4), 450–466. https://doi.org/10.1016/S0022-5371(80)90312-6

Davenport, J. S., & Smith, S. A. (1965). Effects of Hyphenation, Justification and Type Size on Readability. *Journalism Quarterly*, *42*(3), 382–388. https://doi.org/10.1177/107769906504200303

de Beni, R., Borella, E., & Carretti, B. (2007). Reading Comprehension in Aging: The Role of Working Memory and Metacomprehension. *Aging, Neuropsychology, and Cognition*, *14*(2), 189–212. https://doi.org/10.1080/13825580500229213

de Beni, R., & Palladino, P. (2004). Decline in working memory updating through ageing: Intrusion error analyses. *Memory*, *12*(1), 75–89. https://doi.org/10.1080/09658210244000568

DeDe, G., Caplan, D., Kemtes, K., & 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

Dempsey, J., Liu, Q., & Christianson, K. (2020). Convergent probabilistic cues do not trigger syntactic adaptation: Evidence from self-paced reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(10), 1906–1921. https://doi.org/10.1037/xlm0000881

Djiwandono. P. (2006). CULTURAL BIAS IN LANGUAGE TESTING. *TEFLIN Journal*, *17*, 81–88.

Dole. J, Duffy. G, Roehler. L, & Pearson. P. (1991). Moving From the Old to the New: Research on Reading Comprehension Instruction. *Review of Educational Research*, *61*(2), 239–264.

Drieghe, D., & Brysbaert, M. (2002). Strategic effects in associative priming with words, homophones, and pseudohomophones. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(5), 951–961. https://doi.org/10.1037/0278-7393.28.5.951

Dror, I. E., Schmitz-Williams, I. C., & Smith, W. (2005). Older Adults Use Mental Representations That Reduce Cognitive Load: Mental Rotation Utilizes Holistic Representations and Processing. *Experimental Aging Research*, *31*(4), 409–420. https://doi.org/10.1080/03610730500206725

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

Eikmeier, V., Alex-Ruf, S., Maienborn, C., Schröter, H., & Ulrich, R. (2016). *The Mental Timeline During the Processing of Linguistic Information* (pp. 103–122). https://doi.org/10.1075/hcp.52.06eik

Fabiani, M. (2012). It was the best of times, it was the worst of times: A psychophysiologist's view of cognitive aging. *Psychophysiology*, *49*(3), 283–304. https://doi.org/10.1111/j.1469-8986.2011.01331.x

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., & 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., & Patson, N. D. (2007). The 'Good Enough' Approach to Language Comprehension. *Language and Linguistics Compass*, *1*(1–2), 71–83. https://doi.org/10.1111/j.1749-818X.2007.00007.x

Fiore, F., Borella, E., Mammarella, I. C., & de Beni, R. (2012). Age differences in verbal and visuo-spatial working memory updating: Evidence from analysis of serial position curves. *Memory*, *20*(1), 14–27. https://doi.org/10.1080/09658211.2011.628320

Ford, M. (1995). Two modes of mental representation and problem solution in syllogistic reasoning. *Cognition*, *54*(1), 1–71. https://doi.org/10.1016/0010-0277(94)00625-U

Franz, A., & Garnham, A. (1990a). Mental Models as Representations of Discourse and Text. *Language*, *66*(1), 192. https://doi.org/10.2307/415305

Franz, A., & Garnham, A. (1990b). Mental Models as Representations of Discourse and Text. *Language*, *66*(1), 192. https://doi.org/10.2307/415305

Frazier, L., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition*, *6*(4), 291–325. https://doi.org/10.1016/0010-0277(78)90002-1

Fujita, H., & Cunnings, I. (2021). Reanalysis processes in non-native sentence comprehension. *Bilingualism: Language and Cognition*, *24*(4), 628–641. https://doi.org/10.1017/S1366728921000195

Gerard, L., Zacks, R. T., Hasher, L., & Radvansky, G. A. (1991). Age Deficits in Retrieval: The Fan Effect. *Journal of Gerontology*, *46*(4), P131–P136. https://doi.org/10.1093/geronj/46.4.P131

Gilbert, S. J., & Shallice, T. (2002). Task Switching: A PDP Model. *Cognitive Psychology*, *44*(3), 297–337. https://doi.org/10.1006/cogp.2001.0770

Gilchrist, A. L., Cowan, N., & Naveh-Benjamin, M. (2008). Working memory capacity for spoken sentences decreases with adult ageing: Recall of fewer but not smaller chunks in older adults. *Memory*, *16*(7), 773–787. https://doi.org/10.1080/09658210802261124

Gompel. R.P, Pickering M.J, & Traxler M.J. (2000). *Unrestricted race: A new model of syntactic ambiguity resolution* .

Goodwin, G. P., & Johnson-Laird, P. N. (2005). Reasoning about relations. *Psychological Review*, *112*(2), 468–493. https://doi.org/10.1037/0033-295X.112.2.468

Green, M. J., & Mitchell, D. C. (2006a). Absence of real evidence against competition during syntactic ambiguity resolution. *Journal of Memory and Language*, *55*(1), 1–17. https://doi.org/10.1016/j.jml.2006.03.003

Hamm, V. P., & Hasher, L. (1992). Age and the availability of inferences. *Psychology and Aging*, *7*(1), 56–64. https://doi.org/10.1037/0882-7974.7.1.56

Hasher, L., & Zacks, R. T. (1988). *Working Memory, Comprehension, and Aging: A Review and a New View* (pp. 193–225). https://doi.org/10.1016/S0079-7421(08)60041-9

Havlik, A. M., Carmichael, D. A., & Simner, J. (2015). Do sequence-space synaesthetes have better spatial imagery skills? Yes, but there are individual differences. *Cognitive Processing*, *16*(3), 245–253. https://doi.org/10.1007/s10339-015-0657-1

Head, D., & Isom, M. (2010). Age effects on wayfinding and route learning skills. *Behavioural Brain Research*, *209*(1), 49–58. https://doi.org/10.1016/j.bbr.2010.01.012

Hilton, C., Johnson, A., Slattery, T. J., Miellet, S., & Wiener, J. M. (2021). The impact of cognitive aging on route learning rate and the acquisition of landmark knowledge. *Cognition*, *207*, 104524. https://doi.org/10.1016/j.cognition.2020.104524

Hock, M., & Mellard, D. (2005). Reading Comprehension Strategies for Adult Literacy Outcomes. *Journal of Adolescent & Adult Literacy*, *49*(3), 192–200. https://doi.org/10.1598/JAAL.49.3.3

Hollingshead, A. B., & Zipf, G. K. (1949). Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology. *American Sociological Review*, *14*(6), 822. https://doi.org/10.2307/2086695

Hollingshead, & Zipf. (1950). *Zipf, George K. Human behavior and the principle of least effort. Cambridge, (Mass.): Addison-Wesley, 1949, pp. 573* (Vol. 6). https://doi.org/10.1002/1097-4679(195007)6:3<306::AID-JCLP2270060331>3.0.CO;2-7

https://minds.testable.org/. (2021). *Testable Minds*.

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*, 104288. https://doi.org/10.1016/j.jml.2021.104288

Hurlstone, M. J., Hitch, G. J., & Baddeley, A. D. (2014). Memory for serial order across domains: An overview of the literature and directions for future research. *Psychological Bulletin*, *140*(2), 339–373. https://doi.org/10.1037/a0034221

Inhoff, A. W., Kim, A., & Radach, R. (2019). Regressions during Reading. *Vision*, *3*(3), 35. https://doi.org/10.3390/vision3030035

Jahn, G., Knauff, M., & Johnson-Laird, P. N. (2007). Preferred mental models in reasoning about spatial relations. *Memory & Cognition*, *35*(8), 2075–2087. https://doi.org/10.3758/BF03192939

Johnson. R, & Luftig. R. (1982). Encoding of Semantically Distorted Sentences. *The American Journal of Psychology*, *95*(4).

Johnson-Laird. (1991). *Deduction*. Hove and London.

Johnson-Laird, P. N. (2001). Mental models and deduction. *Trends in Cognitive Sciences*, *5*(10), 434–442. https://doi.org/10.1016/S1364-6613(00)01751-4

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*(2), 229–235. https://doi.org/10.1037/a0035948

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

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., Jackson, J. D., Cheung, H., & Anagnopoulos, C. A. (1993). Enhancing older adults' reading comprehension. *Discourse Processes*, *16*(4), 405–428. https://doi.org/10.1080/01638539309544847

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

Keuleers, E., Lacey, P., Rastle, K., & Brysbaert, M. (2012). The British Lexicon Project: Lexical decision data for 28,730 monosyllabic and disyllabic English words. *Behavior Research Methods*, *44*(1), 287–304. https://doi.org/10.3758/s13428-011-0118-4

Kincaid. J, Fishburne. R, Rogers. R, & Chissom.B. (1975). *Derivation Of New Readability Formulas (Automated Readability Index, Fog Count And Flesch Reading Ease Formula) For Navy Enlisted Personnel*.

Kintsch, W., & Keenan, J. (1973). Reading rate and retention as a function of the number of propositions in the base structure of sentences. *Cognitive Psychology*, *5*(3), 257–274. https://doi.org/10.1016/0010-0285(73)90036-4

Klauer, K. C. (1997). Working Memory Involvement in Propositional and Spatial Reasoning. *Thinking & Reasoning*, *3*(1), 9–47. https://doi.org/10.1080/135467897394419

Klencklen, G., Banta Lavenex, P., Brandner, C., & Lavenex, P. (2017). Working memory decline in normal aging: Is it really worse in space than in color? *Learning and Motivation*, *57*, 48–60. https://doi.org/10.1016/j.lmot.2017.01.007

Knauff, M., & May, E. (2006a). Mental imagery, reasoning, and blindness. *Quarterly Journal of Experimental Psychology*, *59*(1), 161–177. https://doi.org/10.1080/17470210500149992

Knauff, M., & May, E. (2006b). Mental imagery, reasoning, and blindness. *Quarterly Journal of Experimental Psychology*, *59*(1), 161–177. https://doi.org/10.1080/17470210500149992

Knauff, M., Strube, G., Jola, C., Rauh, R., & Schlieder, C. (2004). The Psychological Validity of Qualitative Spatial Reasoning in One Dimension. *Spatial Cognition & Computation*, *4*(2), 167–188. https://doi.org/10.1207/s15427633scc0402_3

Krendl, A. C., Kennedy, D. P., Hugenberg, K., & Perry, B. L. (2022). Social Cognitive Abilities Predict Unique Aspects of Older Adults' Personal Social Networks. *The Journals of Gerontology: Series B*, *77*(1), 18–28. https://doi.org/10.1093/geronb/gbab048

Larkin, W., & Burns, D. (1977). Sentence comprehension and memory for embedded structure. *Memory & Cognition*, *5*(1), 17–22. https://doi.org/10.3758/BF03209186

Liefooghe, B., Barrouillet, P., Vandierendonck, A. and Camos, V., 2008. Working memory costs of task switching. Journal of Experimental Psychology: Learning, Memory, and Cognition, 34(3), pp.478-494.

Light, J. (1988). Interaction involving individuals using augmentative and alternative communication systems: State of the art and future directions. AAC: Augmentative and Alternative Communication, 4(2), 66–82. https://doi.org/10.1080/07434618812331274657

Long, D. L., & Prat, C. S. (2008). Individual differences in syntactic ambiguity resolution: Readers vary in their use of plausibility information. *Memory & Cognition*, *36*(2), 375–391. https://doi.org/10.3758/MC.36.2.375

Lowder, M. W., Choi, W., & Gordon, P. C. (2013). Word recognition during reading: The interaction between lexical repetition and frequency. *Memory & Cognition*, *41*(5), 738–751. https://doi.org/10.3758/s13421-012-0288-z

Lowenthal, M. F. (1964). Social Isolation and Mental Illness in Old Age. *American Sociological Review*, *29*(1), 54. https://doi.org/10.2307/2094641

Maass, A., Pagani, D., & Berta, E. (2007). How Beautiful is the Goal and How Violent is the Fistfight? Spatial Bias in the Interpretation of Human Behavior. *Social Cognition*, *25*(6), 833–852. https://doi.org/10.1521/soco.2007.25.6.833

MacDonald, M. C., Just, M. A., & Carpenter, P. A. (1992a). Working memory constraints on the processing of syntactic ambiguity. *Cognitive Psychology*, *24*(1), 56–98. https://doi.org/10.1016/0010-0285(92)90003-K

MacDonald, M. C., Just, M. A., & Carpenter, P. A. (1992b). Working memory constraints on the processing of syntactic ambiguity. *Cognitive Psychology*, *24*(1), 56–98. https://doi.org/10.1016/0010-0285(92)90003-K

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

Margolin, S. J., & Snyder, N. (2018). It may not be that difficult the second time around: the effects of rereading on the comprehension and metacomprehension of negated text. *Journal of Research in Reading*, *41*(2), 392–402. https://doi.org/10.1111/1467-9817.12114

Martínez, E., Mollica, F., & Gibson, E. (2022). Poor writing, not specialized concepts, drives processing difficulty in legal language. *Cognition*, *224*, 105070. https://doi.org/10.1016/j.cognition.2022.105070

Masson, M. E. J., Bub, D. N., Woodward, T. S., & Chan, J. C. K. (2003). Modulation of word-reading processes in task switching. *Journal of Experimental Psychology: General*, *132*(3), 400–418. https://doi.org/10.1037/0096-3445.132.3.400

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

Mohamed, M. T., & Clifton, C. (2011a). Processing Temporary Syntactic Ambiguity: The Effect of Contextual Bias. *Quarterly Journal of Experimental Psychology*, *64*(9), 1797–1820. https://doi.org/10.1080/17470218.2011.582127

Mohamed, M. T., & Clifton, C. (2011b). Processing Temporary Syntactic Ambiguity: The Effect of Contextual Bias. *Quarterly Journal of Experimental Psychology*, *64*(9), 1797–1820. https://doi.org/10.1080/17470218.2011.582127

Morrow, D. G., Stine-Morrow, E. A. L., Leirer, V. O., Andrassy, J. M., & Kahn, J. (1997). The Role of Reader Age and Focus of Attention in Creating Situation Models from Narratives. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *52B*(2), P73–P80. https://doi.org/10.1093/geronb/52B.2.P73

Myerson, J., Hale, S., Wagstaff, D., Poon, L. W., & Smith, G. A. (1990). The information-loss model: A mathematical theory of age-related cognitive slowing. *Psychological Review*, *97*(4), 475–487. https://doi.org/10.1037/0033-295X.97.4.475

Nasreddine, Z. S., Phillips, N. A., BÃ©dirian, 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

Naveh-Benjamin, M. (2000). Adult age differences in memory performance: Tests of an associative deficit hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(5), 1170–1187. https://doi.org/10.1037/0278-7393.26.5.1170

Nejasmic, J., Bucher, L., & Knauff, M. (2015). The construction of spatial mental models—A new view on the continuity effect. *Quarterly Journal of Experimental Psychology*, *68*(9), 1794–1812. https://doi.org/10.1080/17470218.2014.991335

Ni, W. (1996). Sidestepping Garden Paths: Assessing the Contributions of Syntax, Semantics and Plausibility in Resolving Ambiguities. *Language and Cognitive Processes*, *11*(3), 283–334. https://doi.org/10.1080/016909696387196

Oberauer, K., Weidenfeld, A., & Hörnig, R. (2006). Working memory capacity and the construction of spatial mental models in comprehension and deductive reasoning. *Quarterly Journal of Experimental Psychology*, *59*(2), 426–447. https://doi.org/10.1080/17470210500151717

Onifer, W., & Swinney, D. A. (1981). Accessing lexical ambiguities during sentence comprehension: Effects of frequency of meaning and contextual bias. *Memory & Cognition*, *9*(3), 225–236. https://doi.org/10.3758/BF03196957

PAPADOPOULOU, D., & CLAHSEN, H. (2006). Ambiguity resolution in sentence processing: the role of lexical and contextual information. *Journal of Linguistics*, *42*(1), 109–138. https://doi.org/10.1017/S0022226705003701

Park, D. C., Lautenschlager, G., Hedden, T., Davidson, N. S., Smith, A. D., & Smith, P. K. (2002). Models of visuospatial and verbal memory across the adult life span. *Psychology and Aging*, *17*(2), 299–320. https://doi.org/10.1037/0882-7974.17.2.299

Pavlovia.org. (n.d.). *Pavlovia*.

Piantadosi, S. T., Tily, H., & Gibson, E. (2012). The communicative function of ambiguity in language. *Cognition*, *122*(3), 280–291. https://doi.org/10.1016/j.cognition.2011.10.004

Pickering, M. J., & Frisson, S. (2001). Processing ambiguous verbs: Evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(2), 556–573. https://doi.org/10.1037/0278-7393.27.2.556

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

Pliatsikas, C., Veríssimo, J., Babcock, L., Pullman, M. Y., Glei, D. A., Weinstein, M., Goldman, N., & Ullman, M. T. (2019). Working memory in older adults declines with age, but is modulated by sex and education. *Quarterly Journal of Experimental Psychology*, *72*(6), 1308–1327. https://doi.org/10.1177/1747021818791994

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

Preface. (1994). In *The Psychology of Proof* (pp. 32–33). The MIT Press. https://doi.org/10.7551/mitpress/5680.003.0001

Prolific.co. (2021). *Prolific* .

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

Ragni, M., Fangmeier, T., Webber, L., Knauff, M. (2007). Preferred Mental Models: How and Why They Are So Important in Human Reasoning with Spatial Relations. In: Barkowsky, T., Knauff, M., Ligozat, G., Montello, D.R. (eds) Spatial Cognition V Reasoning, Action, Interaction. Spatial Cognition 2006. Lecture Notes in Computer Science(), vol 4387. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-75666-8_11

Rauh, R., Hagen, C., Knauff, M., Kuss, T., & Strube, G. (2005). Preferred and Alternative Mental Models in Spatial Reasoning . *Spatial Cognition and Computation* , *5*(2–3), 239–269.

Rawson, K. A., Dunlosky, J., & Thiede, K. W. (2000a). The rereading effect: Metacomprehension accuracy improves across reading trials. *Memory & Cognition*, *28*(6), 1004–1010. https://doi.org/10.3758/BF03209348

Rawson, K. A., Dunlosky, J., & Thiede, K. W. (2000b). The rereading effect: Metacomprehension accuracy improves across reading trials. *Memory & Cognition*, *28*(6), 1004–1010. https://doi.org/10.3758/BF03209348

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

Rayner, K., Cook, A. E., Juhasz, B. J., & Frazier, L. (2006). Immediate disambiguation of lexically ambiguous words during reading: Evidence from eye movements. *British Journal of Psychology*, *97*(4), 467–482. https://doi.org/10.1348/000712605X89363

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

Rizza, A., & Price, M. C. (2012). Do sequence-space synaesthetes have better spatial imagery skills? Maybe not. *Cognitive Processing*, *13*(S1), 299–303. https://doi.org/10.1007/s10339-012-0459-7

RODDER, W., & KULMANN, F. (2006). Recall and Reasoning—an information theoretical model of cognitive processes. *Information Sciences*, *176*(17), 2439–2466. https://doi.org/10.1016/j.ins.2005.10.010

Román, A., el Fathi, A., & Santiago, J. (2013). Spatial biases in understanding descriptions of static scenes: the role of reading and writing direction. *Memory & Cognition*, *41*(4), 588–599. https://doi.org/10.3758/s13421-012-0285-2

Román, A., Flumini, A., Lizano, P., Escobar, M., & Santiago, J. (2016). Reading direction causes spatial biases in mental model construction in language understanding. *Scientific Reports*, *5*(1), 18248. https://doi.org/10.1038/srep18248

Sachs, J. S. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. *Perception & Psychophysics*, *2*(9), 437–442. https://doi.org/10.3758/BF03208784

Salhi, L., & Bergström, Z. M. (2020). Intact strategic retrieval processes in older adults: no evidence for age-related deficits in source-constrained retrieval. *Memory*, *28*(3), 348–361. https://doi.org/10.1080/09658211.2020.1719161

Schaeken, W., de Vooght, G., Vandierendonck, A., d'Ydewalle, G., & d'Ydewalle, G. (Eds.). (1999a). *Deductive Reasoning and Strategies*. Routledge. https://doi.org/10.4324/9781410603869

Schaeken, W., de Vooght, G., Vandierendonck, A., d'Ydewalle, G., & d'Ydewalle, G. (Eds.). (1999b). *Deductive Reasoning and Strategies*. Routledge. https://doi.org/10.4324/9781410603869

Schendan, H. E. (2017). Implicit Memory☆. In *Reference Module in Neuroscience and Biobehavioral Psychology*. Elsevier. https://doi.org/10.1016/B978-0-12-809324-5.06459-2

Schumaker, J. B., Deshler, D. D., Alley, G. R., Warner, M. M., & Denton, P. H. (1982). *Multipass* : A Learning Strategy for Improving Reading Comprehension. *Learning Disability Quarterly*, *5*(3), 295–304. https://doi.org/10.2307/1510296

Schurer, T., Opitz, B., & Schubert, T. (2020). Working Memory Capacity but Not Prior Knowledge Impact on Readers' Attention and Text Comprehension. *Frontiers in Education*, *5*. https://doi.org/10.3389/feduc.2020.00026

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

Spivey, M. J., & Tanenhaus, M. K. (1998). Syntactic ambiguity resolution in discourse: Modeling the effects of referential context and lexical frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(6), 1521–1543. https://doi.org/10.1037/0278-7393.24.6.1521

Stephen, D. G., Boncoddo, R. A., Magnuson, J. S., & Dixon, J. A. (2009). The dynamics of insight: Mathematical discovery as a phase transition. *Memory & Cognition*, *37*(8), 1132–1149. https://doi.org/10.3758/MC.37.8.1132

Stine-Morrow, E. A. L., Gagne, D. D., Morrow, D. G., & DeWall, B. H. (2004). Age differences in rereading. *Memory & Cognition*, *32*(5), 696–710. https://doi.org/10.3758/BF03195860

Swets, B., Desmet, T., Clifton, C., & Ferreira, F. (2008). Underspecification of syntactic ambiguities: Evidence from self-paced reading. *Memory & Cognition*, *36*(1), 201–216. https://doi.org/10.3758/MC.36.1.201

Swets, B., Desmet, T., Hambrick, D. Z., & Ferreira, F. (2007). The role of working memory in syntactic ambiguity resolution: A psychometric approach. *Journal of Experimental Psychology: General*, *136*(1), 64–81. https://doi.org/10.1037/0096-3445.136.1.64

Swinney, D. A. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior*, *18*(6), 645–659. https://doi.org/10.1016/S0022-5371(79)90355-4

Tabor, W., & Tanenhaus, M. K. (1999). Dynamical Models of Sentence Processing. *Cognitive Science*, *23*(4), 491–515. https://doi.org/10.1207/s15516709cog2304_5

Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic Influences On Parsing: Use of Thematic Role Information in Syntactic Ambiguity Resolution. *Journal of Memory and Language*, *33*(3), 285–318. https://doi.org/10.1006/jmla.1994.1014

Uekermann, J., Thoma, P., & Daum, I. (2008). Proverb interpretation changes in aging. *Brain and Cognition*, *67*(1), 51–57. https://doi.org/10.1016/j.bandc.2007.11.003

Ulrich, R., & Maienborn, C. (2010). Left–right coding of past and future in language: The mental timeline during sentence processing. *Cognition*, *117*(2), 126–138. https://doi.org/10.1016/j.cognition.2010.08.001

Unsworth, N. and Engle, R., 2008. Speed and accuracy of accessing information in working memory: An individual differences investigation of focus switching. Journal of Experimental Psychology: Learning, Memory, and Cognition, 34(3), pp.616-630.

van der Linden, M., Hupet, M., Feyereisen, P., Schelstraete, M.-A., Bestgen, Y., Bruyer, R., Lories, G., el Ahmadi, A., & Seron, X. (1999). Cognitive Mediators of Age-Related Differences in Language Comprehension and Verbal Memory Performance. *Aging, Neuropsychology, and Cognition*, *6*(1), 32–55. https://doi.org/10.1076/anec.6.1.32.791

van Gompel, R. P. G., Pickering, M. J., Pearson, J., & Liversedge, S. P. (2005). Evidence against competition during syntactic ambiguity resolution. *Journal of Memory and Language*, *52*(2), 284–307. https://doi.org/10.1016/j.jml.2004.11.003

van Gompel, R. P. G., Pickering, M. J., & Traxler, M. J. (2000). Unrestricted Race: A New Model of Syntactic Ambiguity Resolution. In *Reading as a Perceptual Process* (pp. 621–648). Elsevier. https://doi.org/10.1016/B978-008043642-5/50029-2

Vandierendonck, A., 2021. How the Working Memory with Distributed Executive Control Model Accounts for Task Switching and Dual-Task Coordination Costs. Journal of Cognition, 4(1).

Vosgerau, Gottfried (2006). The Perceptual Nature of Mental Models, Advances in Psychology 255-275,18 , https://doi.org/10.1016/S0166-4115(06)80039-7.

Vosse, T., & Kempen. G. (2008). Parsing Verb-Final Clauses in German: Garden-path and ERP Effects Modeled by a Parallel Dynamic Parser. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *30*.

Vosse, T., & Kempen, G. (2009). In Defense of Competition During Syntactic Ambiguity Resolution. *Journal of Psycholinguistic Research*, *38*(1), 1–9. https://doi.org/10.1007/s10936-008-9075-1

Wagner, A. D., & Koutstaal, W. (2002). Priming. In *Encyclopedia of the Human Brain* (pp. 27–46). Elsevier. https://doi.org/10.1016/B0-12-227210-2/00286-7

Wallot, S., Lee, J. T., & Kelty-Stephen, D. G. (2019). Switching between reading tasks leads to phase-transitions in reading times in L1 and L2 readers. *PLOS ONE*, *14*(2), e0211502. https://doi.org/10.1371/journal.pone.0211502

Ward, J., Ipser, A., Phanvanova, E., Brown, P., Bunte, I., & Simner, J. (2018). The prevalence and cognitive profile of sequence-space synaesthesia. *Consciousness and Cognition*, *61*, 79–93. https://doi.org/10.1016/j.concog.2018.03.012

Warner, J., & Glass, A. L. (1987). Context and distance-to-disambiguation effects in ambiguity resolution: Evidence from grammaticality judgments of garden path sentences. *Journal of Memory and Language*, *26*(6), 714–738. https://doi.org/10.1016/0749-596X(87)90111-2

Wasow .T, Perfors. A, & Beaver .D. (2005). *The puzzle of ambiguity. Morphology and The Web of Grammar: Essays in Memory of Steven G. Lapointe*. CSLI Publications.

Waters, G. S., & Caplan, D. (2003). The reliability and stability of verbal working memory measures. *Behavior Research Methods, Instruments, & Computers*, *35*(4), 550–564. https://doi.org/10.3758/BF03195534

Wiener, J. M., de Condappa, O., Harris, M. A., & Wolbers, T. (2013). Maladaptive Bias for Extrahippocampal Navigation Strategies in Aging Humans. *Journal of Neuroscience*, *33*(14), 6012–6017. https://doi.org/10.1523/JNEUROSCI.0717-12.2013

Wong, A. Y., & Moss, J. (2022a). Metacomprehension and regressions during reading and rereading. *Metacognition and Learning*, *17*(1), 53–71. https://doi.org/10.1007/s11409-021-09272-w

Wong, A. Y., & Moss, J. (2022b). Metacomprehension and regressions during reading and rereading. *Metacognition and Learning*, *17*(1), 53–71. https://doi.org/10.1007/s11409-021-09272-w

Wood, S. N. (2003). Thin plate regression splines. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *65*(1), 95–114. https://doi.org/10.1111/1467-9868.00374

Wood, S. N. (2004). Stable and Efficient Multiple Smoothing Parameter Estimation for Generalized Additive Models. *Journal of the American Statistical Association*, *99*(467), 673–686. https://doi.org/10.1198/016214504000000980

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *73*(1), 3–36. https://doi.org/10.1111/j.1467-9868.2010.00749.x

Wood, S. N. (2017). *Generalized Additive Models*. Chapman and Hall/CRC. https://doi.org/10.1201/9781315370279

Wood, S. N., Pya, N., & Säfken, B. (2016). Smoothing Parameter and Model Selection for General Smooth Models. *Journal of the American Statistical Association*, *111*(516), 1548–1563. https://doi.org/10.1080/01621459.2016.1180986

Yoo, H., & Dickey, M. W. (2017a). Aging Effects and Working Memory in Garden-Path Recovery. *Clinical Archives of Communication Disorders*, *2*(2), 91–102. https://doi.org/10.21849/cacd.2017.00122

Zwaan, R. A., Magliano, J. P., & Graesser, A. C. (1995). Dimensions of situation model construction in narrative comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(2), 386–397. https://doi.org/10.1037/0278-7393.21.2.386

Appendix:

The following pages contain the appendices for all empirical chapters. Included are all stimulus items The full reproducible codes, raw data, analysis scripts, tables and figures. for each chapter can be found on the OSF link for this thesis https://osf.io/rydeh/?view_only=b75558090201482da9f1ac27dc0fa023
This also contains the anonymised raw data, cleaning protocols and cleaning scripts.

Chapter 2:

Items:

Item	Ambigous Non-Spatial	Unambiguoius non-spatial	Ambiguous Spatial	Unambiguous Spatial
1	The old chapel relied on donations from the local community. The Alder family had donated a great deal of money to the chapel and always considered themselves better because of it. However, it still wasn't enough to match the amount of money that the Dwight family had managed to raise for the chapel. This rivalry had been going on for years and had started to fray relations between everyone in the village. However, a new property developer that had just moved in wanted to grease the wheels of his business plans and so started to spend generously in the village. When Mr. Roberts donated to the chapel, he made sure he at least donated more than the Alder Family.	The old chapel relied on donations from the local community. The Alder family had donated a great deal of money to the chapel and always considered themselves better because of it. However, it still wasn't enough to match the amount of money that the Dwight family had managed to raise for the chapel. This rivalry had been going on for years and had started to fray relations between everyone in the village. However, a new property developer that had just moved in wanted to grease the wheels of his business plans and so started to spend generously in the village. When Mr.Roberts donated to the chapel he made sure he donated less than the Alder Family, so as not to ruffle any feathers.	The old chapel doors stood wide open allowing full view of its interior. Inside there stood a magnificent alter that had been in the church for hundreds of years. To the left of the alter stood the pulpit where the pastor would often hold his sermons from. A recent Archaeological discovery has lead researchers to believe that there is a secret underground chamber somewhere to the right of the pulpit in the church. The local community considered the church an integral part of its identity, though the number of people attending had dwindled significantly over the past few years.	The old chapel doors stood wide open allowing full view of its interior. Inside there stood a magnificent alter that had been in the church for hundreds of years. To the left of the alter stood the pulpit where the pastor would often hold his sermons from. A recent Archaeological discovery has lead researchers to believe that there is a secret underground chamber somewhere to the left of the alter in the church. The local community considered the church an integral part of its identity, though the number of people attending had dwindled significantly over the past few years.
2	Amy was a highly competent hiker, she loved to compete in time trials but was annoyed at the fact that she hadn't managed to climb to the mountain's summit faster than Sam. The two were friends and their rivalry would spur them on to climbing faster and more dangerously in order to one up each other. Amy's friend Bradly wanted to impress her. When he climbed the mountain he made sure that he was at least faster than Amy. The three loved to climb and loved nature so they always made sure to pick up their rubbish wherever they went.	Amy was a highly competent hiker, she loved to compete in time trials but was annoyed at the fact that she hadn't managed to climb to the mountain's summit faster than Sam. The two were friends and their rivalry would spur them on to climbing faster and more dangerously in order to one up each other. Amy's friend Bradly wanted to impress her. When he climbed the mountain but he simply wasn't as fast as Amy. The three loved to climb and loved nature so they always made sure to pick up their rubbish wherever they went.	The mountain stood proudly as it had done for thousands of years. People had always wanted to climb it so over the years a natural rest stop had developed above the camp at the mountain's base. The rest stop wasn't an official one and had over the years become popular with bird watchers. They had no interest in reaching the summit as it was the birds they came for. With a recent increase in tourism on the mountain there was a rumour that somewhere above the camp at the mountain's base there was a small area where people	The mountain stood proudly as it had done for thousands of years. People had always wanted to climb it so over the years a natural rest stop had developed above the camp at the mountain's base. The rest stop wasn't an official one and had over the years become popular with bird watchers. They had no interest in reaching the summit as it was the birds they came for. With a recent increase in tourism on the mountain there was a rumour that somewhere below the camp at the mountain's base there was a small area where people

			would go to throw their rubbish away. The mountain rangers authority found this to be completely unacceptable.	would go to throw their rubbish away. The mountain rangers authority found this to be completely unacceptable.
3	The town surgery had three resident GPs. Malcom who had been at the surgery the longest was liked well enough by the people who came to the surgery but his popularity was overshadowed by Sue who was very well liked by the town's people. The two were old friends so it didn't bother them as to who was more popular. However,with the surgery's refurbishment came an additional GP, Matthew, who's kind nature made him very popular with the towns folk, far more popular than at least Malcom. None of this mattered to the three though as their main goal was to provide the best medical care they could to help their community.	The town surgery had three resident GPs. Malcom who had been at the surgery the longest was liked well enough by the people who came to the surgery but his popularity was overshadowed by Sue who was very well liked by the town's people. The two were old friends so it didn't bother them as to who was more popular. However,with the surgery's refurbishment came an additional GP, Matthew, who's kind nature made him very popular with the towns folk, but not as popular as Malcom. None of this mattered to the three though as their main goal was to provide the best medical care they could to help their community.	The newly refurbished town surgery was finally ready to reopen after a whole month of being closed. The surgery was located in front of a small play park surrounded by a large car park that was used for local shopping. The surgery boasted a great deal more space than it had before, several smaller buildings around were now used to house non medical tools such as cleaning equipment. The cleaner would keep his floor polisher in a shed somewhere behind the surgery. The cleaner had worked at the surgery for as long as anyone could remember and he himself was also quite old.	The newly refurbished town surgery was finally ready to reopen after a whole month of being closed. The surgery was located in front of a small play park surrounded by a large car park that was used for local shopping. The surgery boasted a great deal more space than it had before, Several smaller buildings around were now used to house non medical tools such as cleaning equipment. The cleaner would keep his floor polisher in a shed somewhere in front of the surgery. The cleaner had worked at the surgery for as long as anyone could remember and he himself was also quite old.
4	Benjamin, the mechanic, had only recently arrived on the island. He had been called there on the behest of his friend Allan, who was also a mechanic, in order to fix the islands lighthouse. Benjamin was not as good of a mechanic as Allan but they were sure to make a good team nonetheless. After a long day of trying to repair the lighthouse there was still much work to do to get it working. . Allan decided to call for help. Susan, who also lived on the Island, was not as good of a mechanic as Allan but surely the three of them would complete the work faster than just the two of them.	Benjamin, the mechanic, had only recently arrived on the island. He had been called there on the behest of his friend Allan, who was also a mechanic, in order to fix the islands lighthouse. Benjamin was not as good of a mechanic as Allan but they were sure to make a good team nonetheless. After a long day of trying to repair the lighthouse there was still much work to do to get it working. Allan decided to call for help. Susan, who also lived on the Island, was a better mechanic than Allan and surely the three of them would complete the work faster than just the two of them.	Spice Island lay just three nautical miles off the coast Newfoundland. The island didn't have a great deal to offer in terms of comfort but there were a handful of people that called it home. There was a small weather station to the left the small docks used mostly to load and unload supplies as well as passengers. There was also a small supply store somewhere to the left of the docks which sold various bits and pieces. It was overpriced due to the fact that spice islanders had to import everything from the mainland. Many of them didn't mind spending more on supplies if it meant avoiding a long trip to the mainland.	Spice Island lay just three nautical miles off the coast Newfoundland. The island didn't have a great deal to offer in terms of comfort but there were a handful of people that called it home. There was a small weather station to the left the small docks used mostly to load and unload supplies as well as passengers. There was also a small supply store somewhere to the right of the docks which sold various bits and pieces. It was overpriced due to the fact that spice islanders had to import everything from the mainland. Many of them didn't mind spending more on supplies if it meant avoiding a long trip to the mainland.
5	Terrance had loved caving ever since he was introduced to it by his mother, Joan, as a young boy. Though she was getting older now she was	Terrance had loved caving ever since he was introduced to it by his mother, Joan, as a young boy. Though she was getting older now she was	The local community had long considered the caves at the bottom of the Lincoln gorge dangerous. However, the tourist industry just kept	The local community had long considered the caves at the bottom of the Lincoln gorge dangerous. However, the tourist industry just kept

	still more experienced than him. The two would often plan family vacations that just happened to be in the vicinity of safe but challenging cave systems. Terrance's father had left before he was born, so when his mother had married his stepfather Stephen, Terrance was excited about the family holidays they could go on now as a trio. Stephen had an adventurous streak and also loved caving, though he was also not as experienced as Joan.	still more experienced than him. The two would often plan family vacations that just happened to be in the vicinity of safe but challenging cave systems. Terrance's father had left before he was born, so when his mother had married his stepfather Stephen, Terrance was excited about the family holidays they could go on now as a trio. Stephen had an adventurous streak and also loved caving, and he a far more experienced caver than Joan.	growing. An adventure holiday company had set up a seasoned cavers activity with a rest stop in a large cave somewhere below the entrance. This was usually the only place tourists got to before turning back as the tunnel that went further down was a very tight squeeze. A recent geological survey of the area had determined that there should be a previously undiscovered cavern below the entrance to the caves.	growing. An adventure holiday company had set up a seasoned cavers activity with a rest stop in a large cave somewhere below the entrance. This was usually the only place tourists got to before turning back as the tunnel that went further down was a very tight squeeze. A recent geological survey of the area had determined that there should be a previously undiscovered cavern above the entrance to the caves.
6	Harriet often travelled the long A67 highway. She had what she considered to be one of the most expensive cars in her neighbourhood, far more expensive than her neighbour Barry's car. The two didn't much like each other and it made her happy that he struggled to purchase his car whereas she could easily afford hers. Harriet's other neighbour Olive had a much nicer and more expensive car than Barry. Harriet liked her neighbour Olive.	Harriet often travelled the long A67 highway. She had what she considered to be one of the most expensive cars in her neighbourhood, far more expensive than her neighbour Barry's car. The two didn't much like each other and it made her happy that he struggled to purchase his car whereas she could easily afford hers. Harriet's other neighbour Olive had an even less expensive car than Barry. Harriet liked her neighbour Olive.	The A67 highway was one of the longest roads in the country. It spanned almost all of it, and ran over the river Stig. Heading to Innsmouth from Ipswich people have to cross the bridge over the river. Somewhere before this there was a rest stop that offered travellers bed and breakfast. It was a popular place and saw a great many visitors, making it a profitable place to own. There had been recent reports of a traffic stop and search operations by the police somewhere before reaching the bridge. that had been set up due to the large amounts of drugs being smuggled in the area.	The A67 highway was one of the longest roads in the country. It spanned almost all of it, and ran over the river Stig. Heading to Innsmouth from Ipswich people have to cross the bridge over the river. Somewhere before this there was a rest stop that offered travellers bed and breakfast. It was a popular place and saw a great many visitors, making it a profitable place to own. There had been recent reports of a traffic stop and search operations by the police somewhere after the rest stop. that had been set up due to the large amounts of drugs being smuggled in the area.
7	Sunnydale high school was like any other high school. It was filled with tightly knit groups of friends bound together by their favourite activities. The basketball group was quite popular even more popular than the football group. This may have been down to their recent win streak against their rival high school, which had put them well on their way to the regional basketball finals. On another hand the art club had recently also exploded in popularity due to their incredibly successful summer exhibition which drew crowds from all over	Sunnydale high school was like any other high school. It was filled with tightly knit groups of friends bound together by their favourite activities. The basketball group was quite popular even more popular than the football group. This may have been down to their recent win streak against their rival high school, which had put them well on their way to the regional basketball finals. On another hand the art club had recently also exploded in popularity due to their incredibly successful summer exhibition which drew crowds from all over	The party hall had been rented out for the high school's prom. The theme was 1950's and the hall had been decorated accordingly with bunting and posters from the era. To the right of the table with snacks and refreshments stood a photo booth, where students could have their pictures taken in their fancy dress consumes. Somewhere to the left of the photo booth stood a table where students could write their well wishes to their fellow students, teachers, parents or anyone that had helped them get through the year. This was set to be the	The party hall had been rented out for the high school's prom. The theme was 1950's and the hall had been decorated accordingly with bunting and posters from the era. To the right of the table with snacks and refreshments stood a photo booth, where students could have their pictures taken in their fancy dress consumes. To the right of the photo booth stood a table where students could write their well wishes to their fellow students, teachers, parents or anyone that had helped them get through the year. This was

	town. Though they were still less popular than the basketball group . The groups weren't technically rival groups, but their popularity would influence where the school board would allocate the most funding for the next year.	town. It was apparent that their popularity had surpassed that of the basketball group . The groups weren't technically rival groups, but their popularity would influence where the school board would allocate the most funding for the next year.	biggest event in the school calendar.	set to be the biggest event in the school calendar.
8	The Dunwich marathon drew crowds and runners from all over the county. Mike and his two friends, Ralph and Dustin, had come from two towns over in order to participate. Mike knew that he wasn't as good of a runner as Ralph and Ralph would often be finished for at least ten minutes by the time Mike got to the finish. Dustin was a relatively new friend of theirs having only joined their company a few months ago, but they absolutely loved his company. As soon as the marathon began it became apparent that Dustin, like Ralph was a much better runner than Mike. This was going to be another tough one for Mike.	The Dunwich marathon drew crowds and runners from all over the county. Mike and his two friends, Ralph and Dustin, had come from two towns over in order to participate. Mike knew that he wasn't as good of a runner as Ralph and Ralph would often be finished for at least ten minutes by the time Mike got to the finish. Dustin was a relatively new friend of theirs having only joined their company a few months ago, but they absolutely loved his company. As soon as the marathon began it became apparent that Dustin, unlike Ralph wasn't a better runner than Mike . This was still going to be another tough one for Mike.	The marathon was held every year in Dunwich. Every runner had to have a charity for which they were running. Most of the time the runners that ran for a big charity would run in a group together in order to stand out more for the crowds that gathered along the roads. The Runners of the local hospice knew they were behind the runners of the wildlife charity. It didn't matter who won of course so long as money was raised and everyone had fun. The runners from the cancer charity knew they were infront of the runners from the hospice, which encouraged them. It would be a spectacular finish with a great deal of money raised for charity.	The marathon was held every year in Dunwich. Every runner had to have a charity for which they were running. Most of the time the runners that ran for a big charity would run in a group together in order to stand out more for the crowds that gathered along the roads. The Runners of the local hospice knew they were behind the runners of the wildlife charity. It didn't matter who won of course so long as money was raised and everyone had fun. The runners from the cancer charity knew they were behind of the runners from the hospice, which encouraged them. It would be a spectacular finish with a great deal of money raised for charity.
9	Jonathan had been looking forward to his skiing trip for a while now. Recent profitable returns on some risky investments had made him a lot of money and it was time to flaunt it. He knew that his university rival, Thom, would be skiing at the same resort as him and it made him very happy to know that his net worth was now, way above Thom's. Jonathan had invited his friend Catherine to go skiing with him. Catherine was also very wealthy but not as wealthy as Jonathan.	Jonathan had been looking forward to his skiing trip for a while now. Recent profitable returns on some risky investments had made him a lot of money and it was time to flaunt it. He knew that his university rival, Thom, would be skiing at the same resort as him and it made him very happy to know that his net worth was now, way above Thom's. Jonathan had invited his friend Catherine to go skiing with him. Catherine was even wealthier than Jonathan.	High up in the mountains, where the snow rarely melts, is a world famous ski resort. The resort caters almost exclusively to the incredibly wealthy. It has some of the finest hotels and restaurants in the world. The slope everyone at the resort desired to be seen on was above one of the more exclusive ski lodges. People had to take a special pre booked shuttle bus to get to the lodge itself and then wait to be invited to ski on the slope. Though the lodge couldn't actually stop anyone from skiing on the slope they were responsible for its upkeep and safety. A child has gone missing below the start of the slope and hadn't been seen in several hours.	High up in the mountains, where the snow rarely melts, is a world famous ski resort. The resort caters almost exclusively to the incredibly wealthy. It has some of the finest hotels and restaurants in the world. The slope everyone at the resort desired to be seen on was above one of the more exclusive ski lodges. People had to take a special pre booked shuttle bus to get to the lodge itself and then wait to be invited to ski on the slope. Though the lodge couldn't actually stop anyone from skiing on the slope they were responsible for its upkeep and safety. A child has gone missing above the start of the slope and hadn't been seen in several hours.

10	Agatha was a spectacular horticulturalist. She was better even than Sam who owned the flower shop. If there was something that Sam didn't know or wasn't sure about it was Agatha that she would go to. The two had worked together on multiple projects in the past. Recently a new flower shop had opened up in the next town over, it proprietor was none other than Sam's old rival Rachel who was much more knowledgeable than Sam when it came to Flowers.	Agatha was a spectacular horticulturalist. She was better even than Sam who owned the flower shop. If there was something that Sam didn't know or wasn't sure about it was Agatha that she would go to. The two had worked together on multiple projects in the past. Recently a new flower shop had opened up in the next town over, it proprietor was none other than Sam's old rival Rachel who wasn't as knowledgeable as Sam when it came to Flowers.	The flower shop on Wensley High Street was by far the most popular shop in town. People said it had the best variety of plants on offer. To the right of display of pre-cut flowers was an arrangement of green indoor that would change every week. It was one of the reasons why the shop was so popular, people loved the fact that they could return weekly and find something completely new. Recently a local prankster had taken to putting fake plastic plants with price tags somewhere to the right of the cut flowers, much to the dismay of anyone unfortunate enough to accidentally purchase them.	The flower shop on Wensley High Street was by far the most popular shop in town. People said it had the best variety of plants on offer. To the right of display of pre-cut flowers was an arrangement of green indoor that would change every week. It was one of the reasons why the shop was so popular, people loved the fact that they could return weekly and find something completely new. Recently a local prankster had taken to putting fake plastic plants with price tags somewhere to the left of the cut flowers, much to the dismay of anyone unfortunate enough to accidentally purchase them.
11	Peter, Sally and Joanna were best of friends. The three would spend their summer riding around the village on their bikes. They loved to ride around together. Their parents didn't like the three playing together, they were far more concerned as to what people thought about them. Joanna's family wasn't as well respected as Peter's due to Joanna's brother turning to crime a few years ago and going to prison. Sally's family was also not as well respected as Peter's family. However, none of this ever bothered the three at all. No matter what was done or said to them they always found a way to spend the entire summer together.	Peter, Sally and Joanna were best of friends. The three would spend their summer riding around the village on their bikes. They loved to ride around together. Their parents didn't like the three playing together, they were far more concerned as to what people thought about them. Joanna's family wasn't as well respected as Peter's due to Joanna's brother turning to crime a few years ago and going to prison. Though Sally's family was also more respected Peter's family. However, none of this ever bothered the three at all. No matter what was done or said to them they always found a way to spend the entire summer together.	The village of Northport didn't offer that much in terms of regular amusement but it had some of the best camping around, especially in the summer. Behind the rows of summer lodges was a stream where kids from the village would spend their entire summer. They often played with those that had come to Northport for the camping. Many friendships were forged in and around that stream. Behind the summer lodges stood a monument to a dog that had saved his owners life during a bear attack at the cost of its own.	The village of Northport didn't offer that much in terms of regular amusement but it had some of the best camping around, especially in the summer. Behind the rows of summer lodges was a stream where kids from the village would spend their entire summer. They often played with those that had come to Northport for the camping. Many friendships were forged in and around that stream. In front of the summer lodges stood a monument to a dog that had saved his owners life during a bear attack at the cost of its own.
12	Terry was one of the most experienced pilots in Torton, he'd been flying since he was a boy, he was in fact a far more experienced pilot than Josh who had been managing Torton Airfield for several years now. The two got along fantastically and would often go for drinks together at the pub. This year a new pilot had arrived in Torton, Aaron, an ex-stunt	Terry was one of the most experienced pilots in Torton, he'd been flying since he was a boy, he was in fact a far more experienced pilot than Josh who had been managing Torton Airfield for several years now. The two got along fantastically and would often go for drinks together at the pub. This year a new pilot had arrived in Torton, Aaron, an ex-stunt	The Torton airfield was busy this time of year. Many of the hobby pilots had taken their summer holidays at the exact same time and all wanted to fly. The resident helicopter flight instructor was somewhere above the control tower. He liked to tease the people in the control tower with stunts high in the air that looked like they would end in a	The Torton airfield was busy this time of year. Many of the hobby pilots had taken their summer holidays at the exact same time and all wanted to fly. The resident helicopter flight instructor was somewhere above the control tower. He liked to tease the people in the control tower with stunts high in the air that looked like they would end in a

	pilot who became a flight instructor. Jerry had a great deal more experience than Josh.	pilot who became a flight instructor. Jerry not more experience than Josh.	crash. The tower couldn't care less though they were dealing with a landing request from a small biplane that was having a few engine troubles but it was high above them so it had plenty of space, time and speed to be able to make a safe landing.	crash. The tower couldn't care less though they were dealing with a take-off request from a small biplane that was having a few engine troubles that morning. The tower wasn't sure if the mechanic had resolved the issue so at this point the tower could not OK the take off.
13	The office had decided that in after work in the hellishly hot office that they would all meet up for a drink and barbeque at a pool that they could rent for the evening. David was a better cook than Felix and so David thought it best that he would take care of the barbeque. He thought that the second best cook should still be there at the barbeque in order to help him. A recent arrival at the office, Ella, had previously worked in a famous restaurant in town and was considered, while not as good as David, to be a pretty good cook herself.	The office had decided that in after work in the hellishly hot office that they would all meet up for a drink and barbeque at a pool that they could rent for the evening. David was a better cook than Felix and so David thought it best that he would take care of the barbeque. He thought that the second best cook should still be there at the barbeque in order to help him. A recent arrival at the office, Ella, had previously worked in a famous restaurant in town and was considered a much better cook than David.	The office wasn't really much of an office, it was more of a warehouse with desks. Somewhere to the left of the office managers desk stood a water cooler. She had placed it there in the so that she could keep an eye on her underlings so that they wouldn't be tempted to chat away the whole working day. The air conditioner had recently broken making the office hellishly hot. The manager had placed an order for a large fan so that it would be at least somewhat cooler. She'd made sure that it would be placed somewhere to the right of the watercooler so that she could keep an eye on it and control the settings if she needed to.	The office wasn't really much of an office, it was more of a warehouse with desks. Somewhere to the left of the office managers desk stood a water cooler. She had placed it there in the so that she could keep an eye on her underlings so that they wouldn't be tempted to chat away the whole working day. The air conditioner had recently broken making the office hellishly hot. The manager had placed an order for a large fan so that it would be at least somewhat cooler. She'd made sure that it would be placed somewhere to the left of the watercooler so that she could keep an eye on it and control the settings if she needed to.
14	Three custom jobs had come to Daren in one day. This was most unusual, most of the time he struggled to get 3 jobs in one week, let alone in a day. He was going to need his friends to help him. Daren was a much better metal worker than his friend Steve but he needed his help none the less. The two were good friends and the work was sure to go smoothly with them working together. Toby was relatively new to the trade but was a good worker none the less, he obviously wasn't as good as Daren but he would still be a valuable asset while working on the 3 jobs.	Three custom jobs had come to Daren in one day. This was most unusual, most of the time he struggled to get 3 jobs in one week, let alone in a day. He was going to need his friends to help him. Daren was a much better metal worker than his friend Steve but he needed his help none the less. The two were good friends and the work was sure to go smoothly with them working together. Toby was relatively new to the trade but was a good worker none the less, he was obviously as better than Daren and he would be a valuable asset while working on the 3 jobs.	The metal workshop on Yew street had been there in one form or another since the 1500s and as a result had expanded and grown over the many years of operation. In front of a modern lathe stood a desk with a chair that was used as a mini office. This place was more about metal work than paper work but paper work had to be done nonetheless. Workers had recently noticed some small corners of paper had been chewed off which made them suspect that there was a, or perhaps many mice living behind the desk. It couldn't be in front of the desk simply because that concrete had only recently been repaired.	The metal workshop on Yew street had been there in one form or another since the 1500s and as a result had expanded and grown over the many years of operation. In front of a modern lathe stood a desk with a chair that was used as a mini office. This place was more about metal work than paper work but paper work had to be done nonetheless. Workers had recently noticed some small corners of paper had been chewed off which made them suspect that there was a, or perhaps many mice living in front of the desk. It couldn't be behind of the desk simply because that concrete had only recently been repaired.

15	The workers of the copper mine were celebrating their recent pay rises at the local pub. All of them had been awarded a bonus based on their work experience at the mine and their productivity. Darius was happy, while it was true that he hadn't received as big of bonus as Linda, he was still happy that he would be able to pay off a large proportion of his mortgage. Lucy was also happy with her bonus; she had received more than Darius and was already planning the extravagant holiday she would go on.	The workers of the copper mine were celebrating their recent pay rises at the local pub. All of them had been awarded a bonus based on their work experience at the mine and their productivity. Darius was happy, while it was true that he hadn't received as big of bonus as Linda, he was still happy that he would be able to pay off a large proportion of his mortgage. Lucy was also happy with her bonus; she hadn't received more than Darius but was already planning the extravagant holiday she would go on.	The mine had been a source of copper for almost 50 years. It wasn't particularly big, the company that owned it was small and couldn't afford to expand. Instead they enjoyed a steady stream of revenue and a constant influx of highly skilled workers that would use the small mine as a way to up their experience in the industry. Shaft A was currently the most profitable shaft and was situated somewhere below Shaft B which had run dry a few years ago and was now used as a storage area. A recent exploratory drilling somewhere above Shaft A and had revealed a good deposit of a very copper rich ore.	The mine had been a source of copper for almost 50 years. It wasn't particularly big, the company that owned it was small and couldn't afford to expand. Instead they enjoyed a steady stream of revenue and a constant influx of highly skilled workers that would use the small mine as a way to up their experience in the industry. Shaft A was currently the most profitable shaft and was situated somewhere below Shaft B which had run dry a few years ago and was now used as a storage area. A recent exploratory drilling somewhere below Shaft A and had revealed a good deposit of a very copper rich ore.
16	Jolene was busy serving up noodles from her noodle van in the spot that she had had since the food festival began several years ago. She used to be the only one that sold noodles at the festival but this year she had two competitors. Jolene's noodles weren't as expensive as Tommy's noodles which meant that she was still getting a fair amount of business. Her other competitor, Fanella, also was not as expensive as Tommy. Jolene didn't actually mind all that much, the festival was big enough for the three of them.	Jolene was busy serving up noodles from her noodle van in the spot that she had had since the food festival began several years ago. She used to be the only one that sold noodles at the festival but this year she had two competitors. Jolene's noodles weren't as expensive as Tommy's noodles which meant that she was still getting a fair amount of business. Her other competitor, Fanella, had noodles that were more expensive than Tommy's. Jolene didn't actually mind all that much, the festival was big enough for the three of them.	The food and wine festival was held every year on a field between the towns of Swampscott and Lynn. It was growing year on year. To the left of the entrance were several posters for the upcoming events that made the festival special. Along with the usual fare of eating competitions there were sack races, dog shows and live music. Much to the dismay of many attendees, the wine drinking competition would not be held this year. Somewhere to the left of the entrance someone had forgotten their backpack and there were two festival organisers standing next to it discussing whether the backpack constituted a health and safety hazard.	The food and wine festival was held every year on a field between the towns of Swampscott and Lynn. It was growing year on year. To the left of the entrance were several posters for the upcoming events that made the festival special. Along with the usual fare of eating competitions there were sack races, dog shows and live music. Much to the dismay of many attendees, the wine drinking competition would not be held this year. Somewhere to the right of the entrance someone had forgotten their backpack and there were two festival organisers standing next to it discussing whether the backpack constituted a health and safety hazard.
17	The town of Rockport boasted three major events that always drew in large amounts of crowds as well as a few other smaller events that took place there, usually in the summer. The Folk music festival drew in a much larger crowd than the Rockport Grand Prix but both	The town of Rockport boasted three major events that always drew in large amounts of crowds as well as a few other smaller events that took place there, usually in the summer. The Folk music festival drew in a much larger crowd than the Rockport Grand Prix but both	The Rockport Grand Prix was anything but grand. It was a race track for the under 15s and none of the cars could even go over 40 miles per hour. Not that they would even be able to each such a speed as the course featured many twists and turns so the soapbox	The Rockport Grand Prix was anything but grand. It was a race track for the under 15s and none of the cars could even go over 40 miles per hour. Not that they would even be able to each such a speed as the course featured many twists and turns so the soapbox

	were still popular and brought in a lot of money to the town. The Sailing festival was also more popular the Rockport Grand Prix. Rockport was one of the most up and coming towns in the area and the events were beginning to draw in crowds from further and further afield.	were still popular and brought in a lot of money to the town. The Sailing festival was less popular the Rockport Grand Prix. Rockport was one of the most up and coming towns in the area and the events were beginning to draw in crowds from further and further afield.	like cars had to go slowly to avoid crashing. The Blue team was one of the best in Rockport and was currently in front of the Red team. The coaches were happy about this as they had to beat the Red and Green team to qualify for the finals. The Green team were also ahead of the Reds. The Grand Prix was always a popular event in Rockport and would draw in onlookers from miles around.	like cars had to go slowly to avoid crashing. The Blue team was one of the best in Rockport and was currently in front of the Red team. The coaches were happy about this as they had to beat the Red and Green team to qualify for the finals. The Green team were behind of the Reds. The Grand Prix was always a popular event in Rockport and would draw in onlookers from miles around.
18	The Kingsley hills tourism board were looking to organise the funding of a museum to display the archaeological findings of the area. The board were looking for successful local businesses to co-fund the project. Farmer Jack's farm shop was eager to fund the project as this could mean that they could sell some of their produce to the planned café of the museum. However, they were not as wealthy as The Pepper Mills Hotel and so could couldn't offer as much money as them. Another business that was interested in investing in the project was the local bus company. They were sure to gain additional customers taking people to and from the museum but they too were not as wealthy as the Pepper Mills hotel. The tourism board were sure that with everyone's contribution and potential other fund raising activities that they would have all they needed to build the museum.	The Kingsley hills tourism board were looking to organise the funding of a museum to display the archaeological findings of the area. The board were looking for successful local businesses to co-fund the project. Farmer Jack's farm shop was eager to fund the project as this could mean that they could sell some of their produce to the planned café of the museum. However, they were not as wealthy as The Pepper Mills Hotel and so could couldn't offer as much money as them. Another business that was interested in investing in the project was the local bus company. They were sure to gain additional customers taking people to and from the museum and they were even wealthier than The Pepper Mills hotel. The tourism board were sure that with everyone's contribution and potential other fund raising activities that they would have all they needed to build the museum.	The Kingsley hills were deemed an area of great archaeological importance. There was a small rest stop somewhere below a small fully excavated dig site. The site had yielded many interesting finds that shed light on the daily lives of early Anglo-Saxon settlers in England. Somewhere downhill of this site was a small patch of grass that never grew quite as tall as the grass around it. This was thought to be evidence of another point of interest and a potential dig site. The archaeological work in the hills was often suspended in the summer due to the influx of tourists to the hills.	The Kingsley hills were deemed an area of great archaeological importance. There was a small rest stop somewhere below a small fully excavated dig site. The site had yielded many interesting finds that shed light on the daily lives of early Anglo-Saxon settlers in England. Somewhere upnhill of this site was a small patch of grass that never grew quite as tall as the grass around it. This was thought to be evidence of another point of interest and a potential dig site. The archaeological work in the hills was often suspended in the summer due to the influx of tourists to the hills.
19	Alfred McMurphy had managed the pub ever since his father had retired several years ago. It was a family operation and always had been. Frank McMurphy wasn't as popular with the patrons as his brother Alfred but they liked him well enough when he served the drinks and listened to their woes. Susan McMurphy also	Alfred McMurphy had managed the pub ever since his father had retired several years ago. It was a family operation and always had been. Frank McMurphy wasn't as popular with the patrons as his brother Alfred but they liked him well enough when he served the drinks and listened to their woes. Susan McMurphy also	McMurphy's Pub was one of the best in Amherst. Due to its popularity, the owners had decided to add additional seating so that they could serve more customers than ever before. To the right of the central row of tables was the bar, which was almost as long as the whole pub itself. There were barstools along	McMurphy's Pub was one of the best in Amherst. Due to its popularity, the owners had decided to add additional seating so that they could serve more customers than ever before. To the right of the central row of tables was the bar, which was almost as long as the whole pub itself. There were barstools along

	worked at the pub and while also not as popular as her brother Alfred but was also a popular character in the community.	worked at the pub and was a lot more popular as her brother Alfred and was also a popular character in the community.	the whole bar and were considered to be the best seats in the house. To the left of the bar was an old jukebox that played records from the 70s and 80s. On a Friday night the place was absolutely full, especially if it was one of the bi-weekly live music nights.	the whole bar and were considered to be the best seats in the house. To the Right of the bar was an old jukebox that played records from the 70s and 80s. On a Friday night the place was absolutely full, especially if it was one of the bi-weekly live music nights.
20	The Fishermen of Rockport would often gather in the Pub by the docks and relax after a day out at sea. Today though they'd all stayed on dry land as the regatta was taking place and no one wanted to get in the way. Todd was a skilled captain and loved to buy a round of drinks for the whole pub, especially if he had made a good haul. Although he wasn't as good as Herman who seemed to have a nose for finding where the fish would be on any given day. All the rivalries out at sea were strictly taboo in the pub and they all got along, so long as no one drank too much. Herman's son Melvin had recently purchased his own fishing boat and would head out with his own crew, though only a boy of 18 he was already a better fisherman than Todd.	The Fishermen of Rockport would often gather in the Pub by the docks and relax after a day out at sea. Today though they'd all stayed on dry land as the regatta was taking place and no one wanted to get in the way. Todd was a skilled captain and loved to buy a round of drinks for the whole pub, especially if he had made a good haul. Although he wasn't as good as Herman who seemed to have a nose for finding where the fish would be on any given day. All the rivalries out at sea were strictly taboo in the pub and they all got along, so long as no one drank too much. Herman's son Melvin had recently purchased his own fishing boat and would head out with his own crew, though only a boy of 18 he still wasn't a better fisherman than Todd	The annual sailing festival had started as it always did with the sailing regatta. Three teams from Rockport had entered the regatta in the hopes that they would get some pay off for the weeks of training they had undertaken. The crew of the the Royal Lady had managed to get a good start but were behind the the King Fisher's crew. The weather was perfect for sailing that day and made for much better conditions than the previous years. The third crew from Rockport on the boat the Fast Kitten were also ahead of the the Royal Lady This was going to be an exciting race and the locals cheered on the crews from Rockport as they sailed past them.	The annual sailing festival had started as it always did with the sailing regatta. Three teams from Rockport had entered the regatta in the hopes that they would get some pay off for the weeks of training they had undertaken. The crew of the the Royal Lady had managed to get a good start but were behind the the King Fisher's crew. The weather was perfect for sailing that day and made for much better conditions than the previous years. The third crew from Rockport on the boat the Fast Kitten were behind of the the Royal Lady This was going to be an exciting race and the locals cheered on the crews from Rockport as they sailed past them.
21	There were are a few teams hoping to find Archaeological evidence of Dogger bank settlements that operate in the north sea. The Norwegians have much more funding than the British and their equipment is proof of this as it's some of the most modern equipment on the market. This didn't mean that anyone was actively working against each other though. The Danish teams also have less funding than the Norwegians but they still share their findings with everyone just as the British do.	There were are a few teams hoping to find Archaeological evidence of Dogger bank settlements that operate in the north sea. The Norwegians have much more funding than the British and their equipment is proof of this as it's some of the most modern equipment on the market. This didn't mean that anyone was actively working against each other though. The Danish team's funding was better than the Norwegians but they still share their findings with everyone just as the British do.	Dogger Bank in the North Sea is thought to be one of the most significant locations for evidence of early Neolithic settlements. There is only one problem; it is very much under water. A team of divers and archaeologists was performing a preliminary study of the area. They had sent down divers and a remote control submarine. The submarine was somewhere above the divers who were looking adjusting to the pressure before they dove lower. They hadn't found much at this point at least nothing that was from the time period they had been hoping for. Somewhere	Dogger Bank in the North Sea is thought to be one of the most significant locations for evidence of early Neolithic settlements. There is only one problem; it is very much under water. A team of divers and archaeologists was performing a preliminary study of the area. They had sent down divers and a remote control submarine. The submarine was somewhere above the divers who were looking adjusting to the pressure before they dove lower. They hadn't found much at this point at least nothing that was from the time period they had been hoping for. Somewhere

			below the submarine there was a basket attached to ropes. If the divers found anything, they could place it in the basket and the item could then be hauled up. For the most part the diving teams had been finding plastic, this was going to be a lot harder than everyone thought.	above the submarine there was a basket attached to ropes. If the divers found anything, they could place it in the basket and the item could then be hauled up. For the most part the diving teams had been finding plastic, this was going to be a lot harder than everyone thought.
22	The catering company tasked with providing food and refreshments for the community pool party was having problems. They had had many of their usual staff call in sick that day and were now desperately scrambling to get staff that had booked that day off to come in. Amanda was had only just returned from her holiday so wasn't as likely to come in as Sammy who was always eager to help out, besides Sammy needed the money. There was of course the option of asking Jess to come in but this would have been her only day off so she also wasn't as likely to come in as Sammy.	The catering company tasked with providing food and refreshments for the community pool party was having problems. They had had many of their usual staff call in sick that day and were now desperately scrambling to get staff that had booked that day off to come in. Amanda was had only just returned from her holiday so wasn't as likely to come in as Sammy who was always eager to help out, besides Sammy needed the money. There was of course the option of asking Jess to come in as this wouldn't have been her only day off so she was more likely to come in as Sammy.	The community pool was having its annual party, the whole neighbourhood was invited and entry was free. It was always a great success. To the right of the pool's ladder was the water filter which had been inspected and cleaned before the event. Staff wanted to be sure that everything was up to scratch. The pool was in a better condition than ever this year due to refurbishments that had finished just before summer, the whole staff were very pleased with the results. Though one thing that was overlooked was a small paving slab that had come loose somewhere to the right of the pool's ladder. Hopefully no one would trip over it.	The community pool was having its annual party, the whole neighbourhood was invited and entry was free. It was always a great success. To the right of the pool's ladder was the water filter which had been inspected and cleaned before the event. Staff wanted to be sure that everything was up to scratch. The pool was in a better condition than ever this year due to refurbishments that had finished just before summer, the whole staff were very pleased with the results. Though one thing that was overlooked was a small paving slab that had come loose somewhere to the left of the pool's ladder. Hopefully no one would trip over it.
23	Three business owners from Innsmouth were looking to support the local communities growing art community. They were planning to invest in the gallery. Fred who owned a small Café by the beach was hoping to get his hands on some driftwood pieces that he could display at his café. He wasn't going to be able to pledge as much Money as Geraldine who was trying to be elected as Mayor. Simon was also looking to pledge a decent amount of money to the gallery was also able to double the amount that Geraldine was planning on giving.	Three business owners from Innsmouth were looking to support the local communities growing art community. They were planning to invest in the gallery. Fred who owned a small Café by the beach was hoping to get his hands on some driftwood pieces that he could display at his café. He wasn't going to be able to pledge as much Money as Geraldine who was trying to be elected as Mayor. Simon was also looking to pledge a decent amount of money to the gallery but was also not really able to match the amount that Geraldine was planning on giving.	The art gallery of Innsmouth was holding a special exhibition for up and coming local artists. With Innsmouth only being a small town word had gotten around quickly and as such almost the entire town was there. The Local artist's work was on display throughout a large hall with visitors walking past paintings and sculptures. A sculpture made of driftwood stood behind a smaller sculpture of a Whale at the centre of the hall and seemed to be generating a lot of interest from the locals. Behind the sculpture of a Whale stood a horrible piece made from washed up plastic bottles, which was not very popular	The art gallery of Innsmouth was holding a special exhibition for up and coming local artists. With Innsmouth only being a small town word had gotten around quickly and as such almost the entire town was there. The Local artist's work was on display throughout a large hall with visitors walking past paintings and sculptures. A sculpture made of driftwood stood behind a smaller sculpture of a Whale at the centre of the hall and seemed to be generating a lot of interest from the locals. In front of the sculpture of a Whale stood a horrible piece made from washed up plastic bottles, which was not very popular

			among the locals. Overall the exhibition was looking to be a grand success.	among the locals. Overall the exhibition was looking to be a grand success.
24	It had been some time since Josephine's and Aaron's family had come together, but this was their wedding so no one was going to miss out on it. The two had been together for years and both came from families where people considered themselves the "Heart and Soul" of the party. Aaron's father Gary was a great storyteller, much more so than Aaron was. Josephine also loved to spin yarns but wasn't as good as Gary. The wedding reception and ceremony was bound to be a battleground of sorts as to who would tell the best story.	It had been some time since Josephine's and Darron's family had come together, but this was their wedding so no one was going to miss out on it. The two had been together for years and both came from families where people considered themselves the "Heart and Soul" of the party. Darron's father Gary was a great story teller, much more so than his Darron. Josephine also loved to spin yarns and was better than Gary. The wedding reception and ceremony was bound to be a "battleground" of sorts as to who would tell the best story.	The cemetery was old and within it many trees grew. They had almost out grown the spire of the chapel they surrounded. The Oaks towered above the yew trees . The trees that grew in the cemetery gave plenty of shade so it was never too hot for those that visited their passed relatives in the summer. The Birch trees also towered above the Yews and when they shed their paper like bark it was often confused with confetti left over from a wedding at the chapel. This was a popular spot for weddings, as the trees and the old stonework provided a beautiful backdrop for photos.	The cemetery was old and within it many trees grew. They had almost out grown the spire of the chapel they surrounded. The Oaks towered above the yew trees. The trees that grew in the cemetery gave plenty of shade so it was never too hot for those that visited their passed relatives in the summer. The Birch trees were not as tall as the Yews but when they shed their paper like bark it was often confused with confetti left over from a wedding at the chapel. This was a popular spot for weddings, as the trees and the old stonework provided a beautiful backdrop for photos.

Questions:

Item	Question
1	The Bishop sent a thank you letter with his seal to the top donor every year. Who is most likely to receive this letter?
2	Every year there is a race to the top of the mountain which allows only the best to compete, who is most likely to qualify?
3	Patients have voted for their favourite GP, who is likely to be the most popular?
4	The least skilled mechanic is sent to pick up some coffee, who is this likely to be?
5	The least experienced caver in Terrance's family is made to pack lunch, who is this likely to be?
6	A burgalar is going to steal the most expensive car in the neighbourhood, who's car is most likely to be stolen?
7	Based on their popularity, which of the groups should have the fewest members?
8	After the marathon the three friends decide to compete against each other, who is most likely to win?
9	Jonathan, Catherine and Thom are the first to arrive at the lodge this season, who is the least wealthy out of all of them?

10 A prize is given every year to the most skilled horticulturalist, who is most likely to win?

11 A new family has moved into town and tells their son to stay away from the least respected family. Which family is this likely to be?

12 Next year Taunton airfield will host a small air show, who is the most experienced pilot ?

13 The Barbeque is about to start, which office workers is the worst cook?

14 The least skilled worker in the workshop was always made to clean at the end of the day, who is this likely to be?

15 A tax investigation is launched that will assess work place bonuses. Who has received the biggest bonus out of Lucy and Belinda?

16 Young festival goers tend to go for the cheapest noodles. Who's noodles are they most likely to eat?

17 Based on their popularity, which festival is likely to have the most attendees?

18 Which local company is likely to put the least amount of money toward tourism board's project?

19 All of the McMurphy's are well liked by their patrons but who is the least popular among the three of them?

20 Who is the most skilled fisherman currently in the Rockport pub?

21 The least well funded team is going to be responsible for washing any artifacts found, which team is this likely to be?

22 If the catering company wants only the most skilled workers for the pool party who is least likely to be asked to help?

23 Which of the three business owners is going to donate the least amount of money to the gallery?

24 The wedding will have many speeches and anecdotes but who is likely to tell the best story?

25 Mr. Roberts was standing in front of the altar facing it. In which direction would he go to find where the secret chamber is thought to be?

26 A mountain ranger is at the rest stop which way does he need to go to find where tourists throw their rubbish?

27 The cleaner is standing in the small play park facing away from the surgery, which way does he need to go to pick up his floor polisher from the shed?

28 Allan is standing in front of the weather station facing it, in which direction does he need to go to get to the shop?

29 In which general direction would Joan need to go to find the undiscovered cavern if she started at the rest stop?

30 Harriet is on the bridge heading for the rest stop, which way would she need to go if she wanted to see the the police check point?

31 Which way would one have to go in order to get from the refreshments table to the table with well wishes?

32 Which of the charity running teams is most likely to finish the marathon first?

33 The Lodge staff are leaving the Lodge to look for the missing child, which way should they head?

34 Sam wants to get rid of the plastic plants. She is standing in front of the succulents in which direction should she look to find the plastic plants?

35 A holiday maker has lost his watch by the monument, he is standing by the stream. Which general direction should he head to find his watch ?

36 The helicopter pilot wants to know if he is flying higher than the biplane. Which aircraft is higher?

37 The new fan has been delivered, where would the manager want it to be placed?

38 Daren has decided to set up a humane trap for the mouse where should he put it?

39 The foreman is in the storage area, which way does he need to go to take a look at the newly discovered rich deposit of ore?

40 Tommy has lost his backpack, he is standing in front of the posters facing them, which way does he need to go to get to the lost and found tent?

41 Based on current position, which team is most likely to win?

42 An archaeologist has gone to the potential dig site and wants to get to the rest stop. Which way does he need to go?

43 Susan is resting by the central row of tables and wants to put her favourite song on the Jukebox, which way does she need to go?

44 Which sailing boat from Rockport is currently ahead of the others?

45 The submarine used to transport finds has malfunctioned, which way do the divers need to swim in order to fix it?

46 A lifeguard is standing in front of the water filter facing the pool. In which direction does he need to look to find the loose paving slab?

47 Fred is standing by the Sculpture made of driftwood, in which general direction does he need to go to get to the sculpture made of plastic bottles?

48 A recent storm has brought down the tallest tree in the cemetary, which type of tree is most likely to have fallen?

Item readability metrics:

Item	Ambiguity	SpaNwords	SocNwords	SpaF-K	SocF-K	SpaMeanWordLength	SocMeanWordLength	SpaMeanWordFrequency	SocMeanWordFrequency
1	y	93	88	10.69687	9.50404	4.627907	4.627907	6.135143	6.135143
2	y	102	110	8.14	6.91453	4.330189	4.330189	5.959	5.959
3	y	92	92	9.414261	10.6227	4.707865	4.707865	6.01473	6.01473
4	y	104	103	6.657647	7.00470	4.156863	4.156863	6.421566	6.421566
5	y	88	100	11.11945	10.0107	4.676768	4.676768	5.954643	5.954643
6	y	92	77	6.807391	8.668	4.933333	4.933333	5.787049	5.787049
7	y	102	86	8.293941	8.38604	5.034884	5.034884	5.786438	5.786438
8	y	103	90	7.583137	6.76803	4.41573	4.41573	6.123478	6.123478
9	y	102	92	7.445429	8.36514	4.362637	4.362637	5.976494	5.976494
10	y	91	85	10.36056	8.21819	4.487805	4.487805	6.077143	6.077143
11	y	91	104	7.262112	7.33181	4.639175	4.639175	6.215	6.215
12	y	88	83	7.200598	9.24158	4.493827	4.493827	6.03129	6.03129
13	y	112	88	6.867477	7.83	4.443182	4.443182	6.061096	6.061096
14	y	104	109	8.711961	6.58369	3.981308	3.981308	6.240814	6.240814
15	y	97	95	9.582105	11.1494	4.37234	4.37234	6.102716	6.102716
16	y	93	97	6.807391	9.30055	4.569892	4.569892	6.054935	6.054935
17	y	97	99	5.556237	6.60673	4.591837	4.591837	6.011351	6.011351
18	y	98	101	8.359592	6.52627	4.6875	4.6875	6.028529	6.028529
19	y	107	88	6.550943	7.70441	4.670588	4.670588	6.063492	6.063492
20	y	107	103	7.417857	7.314	4.05	4.05	6.137073	6.137073
21	y	97	91	9.327371	7.96129	4.842697	4.842697	6.099559	6.099559
22	y	100	89	7.035102	6.01217	4.643678	4.643678	5.889697	5.889697
23	y	94	101	8.218441	9.792	4.48	4.48	6.089412	6.089412

24	y	105	101	7.979762	8.52033	4.412371	4.412371	6.175375	6.175375
25	y	103	84	8.139146	7.64258	4.756098	4.756098	5.895224	5.895224
26	y	102	90	4.991598	9.08090	4.409091	4.409091	6.087083	6.087083
1	n	93	88	10.69687	9.50404	4.616279	4.616279	6.09169	6.09169
2	n	102	108	8.14	6.94115	4.339806	4.339806	5.937326	5.937326
3	n	94	92	9.316468	10.6227	4.651685	4.651685	6.047027	6.047027
4	n	104	100	6.657647	7.05510	4.252525	4.252525	6.397125	6.397125
5	n	88	101	11.11945	10.028	4.66	4.66	5.930588	5.930588
6	n	92	75	6.807391	8.69304	4.972603	4.972603	5.754407	5.754407
7	n	101	89	8.14	9.02196	5.05618	5.05618	5.768267	5.768267
8	n	102	92	7.681931	7.17333	4.5	4.5	6.05058	6.05058
9	n	102	91	7.445429	8.34333	4.4	4.4	5.956579	5.956579
10	n	91	85	10.36056	8.35516	4.481481	4.481481	6.078033	6.078033
11	n	93	103	7.19811	7.21273	4.697917	4.697917	6.19	6.19
12	n	93	81	6.663462	9.1725	4.531646	4.531646	6.029333	6.029333
13	n	112	85	6.867477	7.98205	4.588235	4.588235	5.997714	5.997714
14	n	105	106	8.747	6.51881	4	4	6.240118	6.240118
15	n	97	96	9.582105	11.275	4.404255	4.404255	6.092375	6.092375
16	n	93	99	6.807391	9.31646	4.670213	4.670213	6.019615	6.019615
17	n	96	98	5.4	6.48026	4.597938	4.597938	6.000274	6.000274
18	n	98	100	8.359592	6.77792	4.768421	4.768421	5.955	5.955
19	n	107	88	6.550943	7.56558	4.682353	4.682353	6.062698	6.062698
20	n	105	104	7.281538	7.196	4.06	4.06	6.134074	6.134074
21	n	97	91	9.327371	7.96409	4.886364	4.886364	6.116818	6.116818
22	n	100	88	7.035102	5.87653	4.655172	4.655172	5.891194	5.891194
23	n	96	98	8.098684	9.73682	4.505155	4.505155	6.070854	6.070854
24	n	107	99	7.686495	8.49175	4.447917	4.447917	6.170875	6.170875
25	n	103	82	8.139146	7.62676	4.790123	4.790123	5.876119	5.876119

25	n	102	90	4.991598	9.08090	4.409091	4.409091	6.087083	6.087083

Chapter 4:

Items:

	paragraph
	The old chapel relied on donations from the local community. The Alder family had donated a great deal of money to the old chapel, but it was less than the amount of money that the Dwight family had managed to raise for the chapel. However, Mr. Roberts, a new property developer that had just arrived, wanted to grease the wheels of his business plans. He had been spending generously in the village. When Mr. Roberts donated to the chapel, he donated more than the Alder Family.
	Amy was a highly competent hiker and she loved to compete in time trials. She was annoyed at the fact that she was slower than Sam. The two were friends and their rivalry would spur them on to climbing faster and more dangerously in order to one-up each other. Amy's friend Bradley wanted to climb the mountain too. Bradley climbed the mountain faster than Amy. The three loved to climb and loved nature, so they always made sure to pick up their rubbish wherever they went.
	The town surgery had three resident GP's. Malcolm, who had been at the surgery the longest was liked well enough by the people who came to the surgery but he was less popular than Sue, who was very well-liked by patients. With the surgery's refurbishment came a third GP, Matthew. Matthew was more popular than Malcolm. It didn't matter to them though, as their main goal was to provide the best medical care they could to help their community remain healthy.
	Benjamin the mechanic had only just arrived on the island. He had been called there on the behest of his friend Allan, who was also a mechanic, to fix the island's lighthouse. Benjamin was less skilled than Allan. After a long day of trying to repair the lighthouse, there was still much work to do to get it working. Allan decided to call Susan for help. Susan was less skilled than Allan. Surely the three of them would complete the work faster than just the two of them.
	Terrance had loved caving ever since he was introduced to it by his mother, Joan. Joan was more experienced than Terrance. The two would often plan family vacations that just happened to be in the vicinity of safe but challenging cave systems. Terrance's father had left before he was born, so when his mother had married his stepfather Stephen, Terrance was excited about the family holidays they could go on now as a trio. Stephen had an adventurous streak and also loved caving. Stephen was less experienced than Joan.
	Harriet had lived in the little suburb for just a few short months. She was career-driven, had worked very hard, and had recently been promoted at work. She was now wealthier than her neighbour Barry. Harriet was arrogant and boastful, so she often teased him. Harriet was friendly with her other neighbour Olive and the two would often visit one another, even though Olive was wealthier than Barry. Harriet had a good life and enjoyed washing her car every weekend.

	Sunnydale high school was like any high school. The basketball club was more popular than the football club. This may have been down to their recent win streak against their rival high school, which had put them on their way to the basketball finals. On the other hand, the art club had recently also exploded in popularity due to their very successful summer exhibition. The art club were less popular than the basketball club. Their popularity would influence how many members each group had.
	The Dunwich marathon drew runners from all over the county. Mike and his two friends, Ralph and Dustin, had come from two towns over in order to participate. Mike knew that he was slower than Ralph. Dustin was a relatively new friend of theirs, having only joined their company a few months ago, but they absolutely loved him. As soon as the marathon began it became apparent that Dustin was faster than Mike. The event was sponsored by various local businesses.
	Jonathan had been looking forward to his skiing trip for a while now. Recent profitable returns on some risky investments had made him a lot of money and it was time to flaunt it. He knew that his university rival, Thom, would be skiing at the same resort as him and it made him very happy to know that he was wealthier than Thom. Jonathan had invited his childhood friend Catherine to go skiing with him. Catherine had been born rich but wasn't wealthier than Jonathan.
	Agatha was a spectacular horticulturalist. She was more talented than Sam, who owned the local flower shop. If there was something that Sam didn't know or wasn't sure about, it was Agatha that she would go to. The two had worked together on multiple projects in the past. Recently a new flower shop had opened up in the next town over. Its proprietor was none other than Sam's old school rival Rachel. Rachel was more talented with flowers than Sam.
	Peter, Sally, and Joanna were the best of friends. They loved to play together. Their parents didn't like the three playing together, they were far more concerned as to what people thought about them. Joanna's family was less respected than Peter's due to Joanna's brother turning to crime a few years ago and going to prison. Sally's family was less respected than Peter's family. However, none of this ever bothered the three at all. No matter what, they always found a way to spend the entire summer together.
	Torton Airfield was a small private airfield that catered to hobby pilots. Terry was an avid flyer. He had been flying since he was a teenager. He was more experienced than Josh who had been managing Torton airfield for several years now. The two of them got along fantastically and would often go for drinks together at the local pub. Aaron was an ex-stunt pilot and wanted to become a flight instructor at Torton airfield. Aaron was more experienced than Josh.
	The office workers had decided to organise a barbeque. The local beach would be the perfect place to set up. The only thing left was deciding who would be in charge of the grill. Everyone knew David was a better cook than Felix. It would be a big responsibility to ensure that the fine quality meats which had been ordered were cooked to perfection. A recent arrival at the office, Ella, had previously worked in agriculture and was considered a worse cook than David.
	Three custom jobs had come to Daren's workshop in one day. Most of the time he struggled to get 3 jobs in one week, let alone in a day. He was going to need some help. Daren was a much better metal worker than his friend Steve, but he needed his help, nonetheless. They were friends and the work was sure to go smoothly. Toby was relatively new to the trade but was a good worker and a good friend. Toby was a worse metal worker than Daren.
	The workers of the copper mine were celebrating their recent pay rises at the local pub. All of them had been awarded a bonus based on their work experience at the mine and their productivity. Darius was happy, even though he received less money than Belinda. He was still happy that he would be able to pay off a large portion of his mortgage. Lucy was also happy with her bonus. She had received more money than Darius. She was already planning the extravagant holiday she would go on.
	Jolene was busy serving up noodles from her noodle van in the spot that she had had since the food festival began several years ago. She used to be the only one that sold noodles at the festival, but this year she had two competitors. Jolene's noodles were less expensive than Tommy's noodles. Nearby there was another noodle stand. Fanella's noodles were less expensive than Tommy's. Jolene didn't mind all that much, the festival was big enough for the three of them.
	The town of Rockport hosted three major events that always drew in large amounts of crowds. The folk music festival was more popular than the Rockport Grand Prix. Both were important sources of income in the town. The Sailing festival was more popular than the Rockport Grand Prix, but this didn't matter to the townsfolk. Rockport was one of the most up and coming towns in the area. The yearly events were beginning to draw in crowds from further and further afield.
	The Kingsley hills tourism board was looking to build a new museum to display the archaeological findings of the area. The board was eagerly looking for local businesses to co-fund the project. Farmer Jack's farm shop had donated less money than the Pepper-Mills hotel. The local bus company also wanted to co-fund the project, but they had donated less money than the Pepper-Mills hotel. The tourism board was sure that they would have all they needed to build the museum.

	Alfred McMurphy had managed the pub ever since his father had retired several years ago. It was a family operation and always had been. Frank McMurphy was less popular than his brother Alfred, but the patrons liked him well enough. They were all popular when they served the drinks and listened to patrons' woes. Susan McMurphy also worked at the pub and was less popular than her brother Alfred. It didn't matter of course, to the McMurphy's, family always came first.
	The fishermen of Rockport would stay on dry land when the regatta was taking place as no one wanted to get in the way of the racing yachts. There was only one place they wanted to be instead ,the pub. There were many skilled captains in the local pub that day. Todd was less experienced than Herman. Herman's son Melvin had recently purchased his own fishing boat and would head out with his own crew. Melvin was more experienced than Todd.
	There are a few teams hoping to find archaeological evidence of Dogger bank settlements that operate in the North Sea. The Norwegian team has more funding than the British team. Their equipment is proof of this as it's some of the most modern equipment on the market. This doesn’t mean that anyone was actively working against each other though. As in all sciences, working together is key. The Danish team has less funding than the Norwegian team but they still share their findings with everyone just as the British do.
	The catering company that had been hired to provide food for the big pool party was having problems. Many of their usual staff had called in sick. This meant that the company was now desperately scrambling to get their best staff member to come in to work. Amanda was less skilled than Sammy. Both of them could use the extra money. There was also another option, Jess. Jess was less skilled than Sammy. The catering company boss hoped everything would run smoothly.
	Three business owners from Innsmouth were looking to support the local art community and were planning to invest in the gallery. Fred, who owned a small bistro by the beach, donated less money than Geraldine. Geraldine was trying to be elected as mayor of Innsmouth and so had to look as though she cared about local issues. Simon loved art and so was looking to pledge a decent amount of money to the gallery. He donated less money than Geraldine.
	It had been some time since Josephine's and Aaron's families had come together. The two had been together for years and now they were finally getting married. Aaron's father Gary was more popular than Aaron was. Josephine loved to tell jokes in front of an audience. This made her less popular than Gary. The wedding reception and ceremony was bound to be a battleground of sorts as to whom could give the most rousing toast. Although this wedding would be fun for everyone.
	The old chapel's doors stood wide open allowing full view of its interior. Inside there stood a magnificent, hundred-year-old altar. There was a pulpit to the left of the altar, where the pastor would hold his sermons from. A recent archaeological discovery has led researchers to believe that there is a secret underground chamber somewhere to the left of the altar in the church. The local community considered the church an integral part of its identity, but the number of people attending had dwindled significantly over the past few years.
	The mountain stood proudly as it had done for thousands of years. People had always wanted to climb it. There was a rest stop above the camp at the mountain's base. The rest stop had over the years become popular with bird watchers. They had no interest in reaching the summit as it was the birds they came for. There was a rumour that there was a small area where people would go to throw their rubbish away above the camp at the mountain's base.
	The newly refurbished town surgery was finally ready to reopen after a whole month of being closed. The surgery was located in front of a small play park. The surgery boasted more space than it had before, several smaller buildings around were now used to house non-medical tools like cleaning equipment. The cleaner would keep his floor polisher in a shed behind the surgery. The cleaner had worked at the surgery for as long as anyone could remember.
	Spice Island lay just three miles offshore. The island didn't have a lot to offer in terms of comfort but there were a handful of people that called it home. There was a small weather station to the left the small dock used to load and unload supplies. There was also a supply store to the left of the docks. It was overpriced because everything had to be imported from the mainland. Many islanders didn't mind spending more on supplies if it meant avoiding a long trip to the mainland.
	The locals had long considered the caves at the bottom of the Lincoln gorge dangerous. However, the tourist industry just kept growing. An adventure holiday company had set up here too. There was a rest stop below the largest cavern. The largest cavern was usually the only place inexperienced cavers got to before turning back as the tunnel that went further was a very tight squeeze. A recent geological survey of the area had determined that there should be a previously undiscovered cavern below the largest cavern.
	The A67 highway was the longest road in the county. Along the highway were numerous landmarks that overlooked the scenic river valley. There was a small rest stop and petrol station in front of an old bridge. The rest stop was a popular place due to its great scenic overlook and saw a great many visitors. This made it a profitable place to own. There was also a statue of a local folk hero behind the rest stop.

	The hall had been rented out for the high school's prom. The theme was 1950's and the hall had been decorated accordingly with bunting and posters from the era. There was a photo booth to the right of a table with snacks and refreshments. There was also a pinboard to the left of the photo booth. The pinboard was where students could write inspirational quotes and messages for others to see. All the tickets for the prom had sold out almost immediately.
	The marathon was held every year in Dunwich. All runners ran for a charity of their choosing and would run in a group together. The runners of the local hospice were behind the runners of the wildlife charity. It didn't matter who won, of course, so long as everyone had fun. The runners from the cancer charity were in front of the runners from the hospice. It would be a spectacular finish with a great deal of money raised for charity.
	High up in the mountains, is a world-famous ski resort. The resort caters almost exclusively to the incredibly wealthy. The slope everyone at the resort desired to be seen on was above the exclusive ski lodge. People had to be specially invited to stay at the lodge. Though the lodge couldn't actually stop anyone from skiing on the slope, they were responsible for its upkeep and safety. A child has gone missing below the start of the slope and hadn't been seen in several hours.
	The flower shop on Wensley High Street was the town's favourite. The succulents were displayed to the right of the pre-cut flowers. It was one of the reasons why the shop was so popular, people loved the fact that they could return weekly and find something new. Recently a local prankster had taken to putting fake plastic plants with price tags to the right of the cut flowers, much to the dismay of anyone unfortunate enough to accidentally purchase them.
	The village of Northport didn't offer that much in terms of regular amusement, but it had some of the best camping around. There was a stream behind the rows of summer lodges. Children from the village would spend their entire summer playing there. They often played with those that had come to Northport for their holiday. Many friendships were forged in and around that stream. There was a monument to a dog that had saved his owner's life during a bear attack behind the summer lodges.
	The Torton airfield was busy this time of year. The resident helicopter flight instructor was flying above the control tower. He liked to tease the people in the control tower with stunts that looked like they would end in a crash. The tower couldn't care less though. They were dealing with a small biplane that was having a few engine troubles that morning. The biplane was flying above the control tower. The airfield would always get busy this time of year.
	The office was always busy this time of year. There was a water cooler to the left of the manager's desk. The manager had placed it there so that she could keep an eye on her underlings so that they wouldn't be tempted to chat away the whole working day. The air conditioner had broken making the office hellishly hot. The manager had ordered a new fan to cool the office down. The fan was placed to the right of the water cooler.
	The little metal workshop on Yew street had been there since the 80s. There was a small desk with a chair in front of a modern lathe. This place was more about metal work than paperwork, but the paperwork had to be done, nonetheless. Workers had recently noticed that some small corners of paper had been chewed off which made them suspect that a mouse was living somewhere. They believed that the mouse's hole was behind the desk.
	The mine had been a source of copper for almost 50 years. It wasn't particularly big, the company that owned it was small and couldn't afford to expand. Instead, they enjoyed a steady stream of revenue and a constant influx of highly skilled workers. Shaft-A was currently the most profitable shaft and was situated below Shaft-B which was used as a storage area. There was also an exploratory drill site above Shaft-A. The exploratory drilling had revealed a good copper-rich ore vein.
	The food and wine festival was held every year in the Swampscott town square. It was growing year on year. There were posters for the upcoming events to the left of the entrance. Along with the usual fare of eating competitions, there were sack races, dog shows, and live music. Much to the dismay of many attendees, the wine drinking competition would not be held this year. The tent used to return lost and found items to their owners was to the left of the entrance.
	The Rockport Grand Prix was anything but grand. It was a go-cart race for the under 15s and none of the carts could go above 30 miles per hour. The Blue team was in front of the Red team. The coaches were happy about this as they had to beat the Red and Green teams to qualify for the finals. The Green team was in front of the Red team. The Grand Prix was always a popular event in Rockport and would draw in onlookers from miles around.
	The Kingsley hills were an area of great archaeological importance. There was a small rest stop below a fully excavated dig site. The site had yielded many interesting finds that shed light on the daily lives of early Anglo-Saxon settlers in England. There was also another potential dig site below the fully excavated one. This site would soon undergo a full investigation. The archaeological work in the hills was often suspended in the summer due to the influx of tourists.

	McMurphy's Pub was one of the best in Amherst. The owners had decided to add additional seating so that they could serve more customers than ever before. The bar was to the right of the central row of tables. There were barstools along the whole bar and were considered to be the best seats in the house. There was also an old jukebox to the left of the bar. On a Friday night, the place was full, especially if it was one of the bi-weekly live music nights.
	The annual sailing festival had started as it always did with the sailing regatta. Three teams from Rockport had entered the regatta in the hope of winning. The racing yacht "Royal Lady" was behind the yacht "King Fisher". The weather was perfect for sailing that day and made for much better conditions than the previous years. The yacht "Fast Kitten" was in front of the "Royal Lady". This was going to be an exciting race and the locals cheered on the crews from Rockport as they sailed past them.
	Dogger Bank in the North Sea is thought to be one of the most significant locations for evidence of early Neolithic settlements. There is only one problem, it is underwater. A team of divers and archaeologists were performing a study of the area. The submarine in charge of taking pictures was above the divers who were adjusting to the pressure before they dove lower. There was also a transport submarine below the submarine taking pictures. This would be used to bring anything the divers found to the surface.
	The community pool was having its annual party. The water filter was to the right of the pool's ladder. This had been inspected and cleaned before the event. Staff wanted to be sure that everything was up to scratch. The pool was in better condition than ever this year due to refurbishments that had finished just before summer. Though one thing was overlooked, there was a loose paving slab to the right of the pool's ladder. Hopefully, no one would trip over it.
	The art gallery of Innsmouth was holding an exhibition for local artists. The artist's work was on display throughout a large hall. A sculpture made of driftwood stood behind a smaller sculpture of a whale at the centre of the hall. There was also a sculpture made of washed-up plastic bottles behind the sculpture of the whale. This sculpture was not very popular among the locals. Overall, the exhibition was looking to be a grand success and many pieces would be sold.
	The cemetery was old and within it many trees grew. On the outside wall of the chapel, there were many stone carvings. There was a gargoyle above a memorial plaque. The trees that grew in the cemetery gave plenty of shade, so it was never too hot for those that visited their dead relatives in the summer. There was also a small protective symbol carved into the wall of the chapel above the memorial plaque. The trees and the old stonework provided a beautiful backdrop for wedding photos.

Questions:

item	cond	Q1	Q2
1	1	The Bishop sent a thank you letter to the top donor every year. Who is most likely to receive this letter?	What does the old chapel rely on?
2	1	Every year there is a race to the top of the mountain, which allows only the best to compete. Who is most likely to qualify?	The time trials take place on a _______

3	1	Patients have voted for their favourite GP. Who is likely to be the most popular?	The GP's surgery has recently_____
4	1	The least skilled mechanic is sent to pick up some coffee. Who is this likely to be?	What are the three mechanics trying to fix?
5	1	The least experienced caver in Terrance's family is made to pack lunch. Who is this likely to be?	What does the family like to do?
6	1	A burglar wants to burgle the wealthiest person in the neighbourhood. Whose house should he choose?	What does Harriet enjoy doing every weekend?
7	1	Based on their popularity, which of the groups should have the fewest members?	What is the high school called?
8	1	After the marathon, the three friends decide to compete against each other. Who is most likely to win?	The marathon draws crowds from all over the___
9	1	Jonathan, Catherine, and Thom are the first to arrive at the lodge this season. Who is the least wealthy out of all of them?	What do all the people mentioned have in common?
10	1	A prize is given every year to the most skilled horticulturalist. Who is most likely to win?	What kind of shop does Sam own?
11	1	A new family has moved into town and tells their son to stay away from the least respected family. Which family is this likely to be?	What season do the three friends always find a way to spend together?

12	1	Next year Taunton airfield will host a small air show. Who is the most experienced pilot?	When did Terry start flying?
13	1	The Barbeque is about to start. Which office worker is the worst cook?	What are the employees of the office organising?
14	1	The least skilled worker in the workshop was always made to clean at the end of the day. Who is this likely to be?	How many custom jobs are the crew working on?
15	1	A tax investigation is launched that will assess workplace bonuses. Who has received the biggest bonus?	What have the workers at the mine received?
16	1	Young festival-goers prefer the cheapest noodles. Whose noodles are they most likely to eat?	What does Jolene's van sell?
17	1	Based on their popularity, which festival is likely to have the most attendees?	What is the town called?
18	1	Which local company is likely to put the least amount of money toward the tourism board's project?	What does the tourism board want to build?
19	1	All of the McMurphy's are well-liked by their patrons but who is the least popular among the three of them?	What is the family's last name?
20	1	Who is the most skilled fisherman currently in the Rockport pub?	Why are the fishermen not fishing?
21	1	The least funded team is going to be responsible for washing any artifacts found. Which team is this likely to be?	In what sea is Dogger bank located?

22	1	If the catering company wants only the most skilled workers for the pool party who is least likely to be asked to help?	What event has the catering company been hired to cater for?
23	1	Which of the three business owners is going to donate the least amount of money to the gallery?	What is the town called?
24	1	Who is likely to be the least popular person at the wedding?	What event is taking place?
25	1	Mr. Roberts was standing in front of the altar facing it. In which direction would he go to find where the secret chamber is thought to be?	What has happened to the number of people attending service in the chapel?
26	1	A mountain ranger is at the rest stop. Which way does he need to go to find where tourists throw their rubbish?	The rest stop has become popular with which group?
27	1	The cleaner is standing in the small play park facing away from the surgery. Which way does he need to go to pick up his floor polisher from the shed?	Roughly how long has the cleaner worked at the surgery?
28	1	Allan is standing in front of the weather station facing it. In which direction does he need to go to get to the shop?	The island has a ____ population?
29	1	In which general direction would Joan need to go to find the undiscovered cavern if she started at the rest stop?	What is the gorge called?
30	1	Tom is standing by the bridge facing away from the rest stop. Which way does he need to go to get to the statue?	Roughly how long is the highway?

31	1	Which way would you have to go in order to get from the refreshments table to the pinboard?	What is the theme of the prom?
32	1	Which of the charity running teams is most likely to finish the marathon first?	How often is the marathon held?
33	1	The lodge staff are leaving the lodge to look for the missing child. Which way should they head?	What sort of people does the resort cater to?
34	1	Sam wants to get rid of the plastic plants, she is standing in front of the succulents. In which direction should she look to find the plastic plants?	Where is the flower shop?
35	1	A holidaymaker has lost his watch by the monument, he is standing by the stream with his back to the lodges. Which general direction should he head to find his watch?	What is the village called?
36	1	A photographer wants to know what is highest in the sky. What is it likely to be?	When will the airfield get less busy?
37	1	You are standing at the water cooler facing the same direction as the manager. In which direction is the new fan?	What is the temperature in the office?
38	1	The mouse is sitting by the lathe facing away from the desk. Which way does it need to go to get home?	Roughly when did the workshop first open?
39	1	The foreman is in shaft B. In which direction should he take the elevator to get to the exploratory drill site?	What does the mine produce?

40	1	Tommy has lost his backpack; he is standing in front of the posters facing them. Which way does he need to go to get to the lost and found tent?	Where is the festival being held?
41	1	Based on the current position, which team is most likely to win?	What sort of race is the Rockport Grand Prix?
42	1	An archaeologist has gone to the potential dig site and wants to get to the rest stop. Which way does he need to go?	For what people would the Kingsley hills be important?
43	1	Susan is resting by the central row of tables and wants to put her favourite song on the Jukebox. Which way does she need to go?	Where is the pub located?
44	1	Which sailing boat from Rockport is currently ahead of the others?	How often is the sailing festival held?
45	1	The submarine used to transport finds has malfunctioned. Which way do the divers need to swim in order to fix it?	How many submarines are the archaeologists currently using?
46	1	A lifeguard is standing in front of the water filter facing the pool. In which direction does he need to look to find the loose paving slab?	When did the refurbishments take place?
47	1	Fred is standing by the sculpture made of driftwood, facing away from the whale sculpture. In which general direction does he need to go to get to the sculpture made of plastic bottles?	The gallery is currently displaying art from which sort of artist?
48	1	A spider is sitting on the gargoyle. If it wants to sit on the protective symbol, which way does it need to go?	What gives the cemetery its shade in the summer?

1	2	The Bishop sent a thank you letter to the top donor every year. Who is most likely to receive this letter?	What does the old chapel rely on?
2	2	Every year there is a race to the top of the mountain, which allows only the best to compete. Who is most likely to qualify?	The time trials take place on a _______
3	2	Patients have voted for their favourite GP. Who is likely to be the most popular?	The GP's surgery has recently_____
4	2	The least skilled mechanic is sent to pick up some coffee. Who is this likely to be?	What are the three mechanics trying to fix?
5	2	The least experienced caver in Terrance's family is made to pack lunch. Who is this likely to be?	What does the family like to do?
6	2	A burglar wants to burgle the wealthiest person in the neighbourhood. Whose house should he choose?	What does Harriet enjoy doing every weekend?
7	2	Based on their popularity, which of the groups should have the fewest members?	What is the high school called?
8	2	After the marathon, the three friends decide to compete against each other. Who is most likely to win?	The marathon draws crowds from all over the___
9	2	Jonathan, Catherine, and Thom are the first to arrive at the lodge this season. Who is the least wealthy out of all of them?	What do all the people mentioned have in common?

10	2	A prize is given every year to the most skilled horticulturalist. Who is most likely to win?	What kind of shop does Sam own?
11	2	A new family has moved into town and tells their son to stay away from the least respected family. Which family is this likely to be?	What season do the three friends always find a way to spend together?
12	2	Next year Taunton airfield will host a small air show. Who is the most experienced pilot?	When did Terry start flying?
13	2	The Barbeque is about to start. Which office worker is the worst cook?	What are the employees of the office organising?
14	2	The least skilled worker in the workshop was always made to clean at the end of the day. Who is this likely to be?	How many custom jobs are the crew working on?
15	2	A tax investigation is launched that will assess workplace bonuses. Who has received the biggest bonus?	What have the workers at the mine received?
16	2	Young festival-goers prefer the cheapest noodles. Whose noodles are they most likely to eat?	What does Jolene's van sell?
17	2	Based on their popularity, which festival is likely to have the most attendees?	What is the town called?
18	2	Which local company is likely to put the least amount of money toward the tourism board's project?	What does the tourism board want to build?

19	2	All of the McMurphy's are well-liked by their patrons but who is the least popular among the three of them?	What is the family's last name?
20	2	Who is the most skilled fisherman currently in the Rockport pub?	Why are the fishermen not fishing?
21	2	The least funded team is going to be responsible for washing any artifacts found. Which team is this likely to be?	In what sea is Dogger bank located?
22	2	If the catering company wants only the most skilled workers for the pool party who is least likely to be asked to help?	What event has the catering company been hired to cater for?
23	2	Which of the three business owners is going to donate the least amount of money to the gallery?	What is the town called?
24	2	Who is likely to be the least popular person at the wedding?	What event is taking place?
25	2	Mr. Roberts was standing in front of the altar facing it. In which direction would he go to find where the secret chamber is thought to be?	What has happened to the number of people attending service in the chapel?
26	2	A mountain ranger is at the rest stop. Which way does he need to go to find where tourists throw their rubbish?	The rest stop has become popular with which group?
27	2	The cleaner is standing in the small play park facing away from the surgery. Which way does he need to go to pick up his floor polisher from the shed?	Roughly how long has the cleaner worked at the surgery?

28	2	Allan is standing in front of the weather station facing it. In which direction does he need to go to get to the shop?	The island has a ____ population?
29	2	In which general direction would Joan need to go to find the undiscovered cavern if she started at the rest stop?	What is the gorge called?
30	2	Tom is standing by the bridge facing away from the rest stop. Which way does he need to go to get to the statue?	Roughly how long is the highway?
31	2	Which way would you have to go in order to get from the refreshments table to the pinboard?	What is the theme of the prom?
32	2	Which of the charity running teams is most likely to finish the marathon first?	How often is the marathon held?
33	2	The lodge staff are leaving the lodge to look for the missing child. Which way should they head?	What sort of people does the resort cater to?
34	2	Sam wants to get rid of the plastic plants, she is standing in front of the succulents. In which direction should she look to find the plastic plants?	Where is the flower shop?
35	2	A holidaymaker has lost his watch by the monument, he is standing by the stream with his back to the lodges. Which general direction should he head to find his watch?	What is the village called?
36	2	A photographer wants to know what is highest in the sky. What is it likely to be?	When will the airfield get less busy?

37	2	You are standing at the water cooler facing the same direction as the manager. In which direction is the new fan?	What is the temperature in the office?
38	2	The mouse is sitting by the lathe facing away from the desk. Which way does it need to go to get home?	Roughly when did the workshop first open?
39	2	The foreman is in shaft B. In which direction should he take the elevator to get to the exploratory drill site?	What does the mine produce?
40	2	Tommy has lost his backpack; he is standing in front of the posters facing them. Which way does he need to go to get to the lost and found tent?	Where is the festival being held?
41	2	Based on the current position, which team is most likely to win?	What sort of race is the Rockport Grand Prix?
42	2	An archaeologist has gone to the potential dig site and wants to get to the rest stop. Which way does he need to go?	For what people would the Kingsley hills be important?
43	2	Susan is resting by the central row of tables and wants to put her favourite song on the Jukebox. Which way does she need to go?	Where is the pub located?
44	2	Which sailing boat from Rockport is currently ahead of the others?	How often is the sailing festival held?
45	2	The submarine used to transport finds has malfunctioned. Which way do the divers need to swim in order to fix it?	How many submarines are the archaeologists currently using?

46	2	A lifeguard is standing in front of the water filter facing the pool. In which direction does he need to look to find the loose paving slab?	When did the refurbishments take place?
47	2	Fred is standing by the sculpture made of driftwood, facing away from the whale sculpture. In which general direction does he need to go to get to the sculpture made of plastic bottles?	The gallery is currently displaying art from which sort of artist?
48	2	A spider is sitting on the gargoyle. If it wants to sit on the protective symbol, which way does it need to go?	What gives the cemetery its shade in the summer?

Supplementary eye-tracking data:

		Fixation Duration		Saccade Duration		Saccade Length		Regression Likelihood		Launch Site Launch Site		Landing Position	
oiguity	ST	Old	oung	Old	oung	Old	oung	Old	oung	Old	oung	Old	oung
iguous	atial	(424)	(512)	(267)	(363)	7 (18.)	4 (15)	(0.482)	(0.497)	15.64)	12.69)	(14.86)	(9.73)
iguous	Spatial	(112)	3 (61)	(201)	(614)	9 (17.)	6 (14)	(0.481)	(0.499)	(14.83)	(11.63)	12.211)	(8.30)
biguous	atial	(424)	) (51)	(267)	(363)	7 (18.)	4 (15)	(0.482)	(0.497)	15.64)	12.69)	(14.86)	(9.73)
biguous	Spatial	(112)	3 (61)	(201)	(614)	9 (17.)	6 (14)	(0.481)	(0.499)	(14.83)	(11.63)	(12.21)	(8.30)

Chapter 5:

Items:

item	paragraph
1	The old chapel relied on donations from the local community. The Alder family had donated a great deal of money to the old chapel, but it was less than the amount of money that the Dwight family had managed to raise for the chapel. However, Mr. Roberts, a new property developer that had just arrived, wanted to grease the wheels of his business plans. He had been spending generously in the village. When Mr. Roberts donated to the chapel, he donated more than the Alder Family.
2	Amy was a highly competent hiker and she loved to compete in time trials. She was annoyed at the fact that she was slower than Sam. The two were friends and their rivalry would spur them on to climbing faster and more dangerously in order to one-up each other. Amy's friend Bradley wanted to climb the mountain too. Bradley climbed the mountain slower than Amy. The three loved to climb and loved nature, so they always made sure to pick up their rubbish wherever they went.
3	The town surgery had three resident GP's. Malcolm, who had been at the surgery the longest was liked well enough by the people who came to the surgery but he was less popular than Sue, who was very well-liked by patients. With the surgery's refurbishment came a third GP, Matthew. Matthew was more popular than Malcolm. It didn't matter to them though, as their main goal was to provide the best medical care they could to help their community remain healthy.
4	Benjamin the mechanic had only just arrived on the island. He had been called there on the behest of his friend Allan, who was also a mechanic, to fix the island's lighthouse. Benjamin was less skilled than Allan. After a long day of trying to repair the lighthouse, there was still much work to do to get it working. Allan decided to call Susan for help. Susan was more skilled than Allan. Surely the three of them would complete the work faster than just the two of them.
5	Terrance had loved caving ever since he was introduced to it by his mother, Joan. Joan was more experienced than Terrance. The two would often plan family vacations that just happened to be in the vicinity of safe but challenging cave systems. Terrance's father had left before he was born, so when his mother had married his stepfather Stephen, Terrance was excited about the family holidays they could go on now as a trio. Stephen had an adventurous streak and also loved caving. Stephen was less experienced than Joan.
6	Harriet had lived in the little suburb for just a few short months. She was career-driven, had worked very hard, and had recently been promoted at work. She was now wealthier than her neighbour Barry. Harriet was arrogant and boastful, so she often teased him. Harriet was friendly with her other neighbour Olive and the two would often visit one another, even though Olive wasn't wealthier than Barry. Harriet had a good life and enjoyed washing her car every weekend.
7	Sunnydale high school was like any high school. The basketball club was more popular than the football club. This may have been down to their recent win streak against their rival high school, which had put them on their way to the basketball finals. On the other hand, the art club had recently also exploded in popularity due to their very successful summer exhibition. The art club were less popular than the basketball club. Their popularity would influence how many members each group had.
8	The Dunwich marathon drew runners from all over the county. Mike and his two friends, Ralph and Dustin, had come from two towns over in order to participate. Mike knew that he was slower than Ralph. Dustin was a relatively new friend of theirs, having only joined their company a few months ago, but they absolutely loved him. As soon as the marathon began it became apparent that Dustin was slower than Mike. The event was sponsored by various local businesses.
9	Jonathan had been looking forward to his skiing trip for a while now. Recent profitable returns on some risky investments had made him a lot of money and it was time to flaunt it. He knew that his university rival, Thom, would be skiing at the same resort as him and it made him very happy to know that he was wealthier than

Thom. Jonathan had invited his childhood friend Catherine to go skiing with him. Catherine had been born rich but wasn't wealthier than Jonathan.

10 Agatha was a spectacular horticulturalist. She was more talented than Sam, who owned the local flower shop. If there was something that Sam didn't know or wasn't sure about, it was Agatha that she would go to. The two had worked together on multiple projects in the past. Recently a new flower shop had opened up in the next town over. Its proprietor was none other than Sam's old school rival Rachel. Rachel was less talented with flowers than Sam.

11 Peter, Sally, and Joanna were the best of friends. They loved to play together. Their parents didn't like the three playing together, they were far more concerned as to what people thought about them. Joanna's family was less respected than Peter's due to Joanna's brother turning to crime a few years ago and going to prison. Sally's family was less respected than Peter's family. However, none of this ever bothered the three at all. No matter what, they always found a way to spend the entire summer together.

12 Torton Airfield was a small private airfield that catered to hobby pilots. Terry was an avid flyer. He had been flying since he was a teenager. He was more experienced than Josh who had been managing Torton airfield for several years now. The two of them got along fantastically and would often go for drinks together at the local pub. Aaron was an ex-stunt pilot and wanted to become a flight instructor at Torton airfield. Aaron was less experienced than Josh.

13 The office workers had decided to organise a barbeque. The local beach would be the perfect place to set up. The only thing left was deciding who would be in charge of the grill. Everyone knew David was a better cook than Felix. It would be a big responsibility to ensure that the fine quality meats which had been ordered were cooked to perfection. A recent arrival at the office, Ella, had previously worked in agriculture and was considered a worse cook than David.

14 Three custom jobs had come to Daren's workshop in one day. Most of the time he struggled to get 3 jobs in one week, let alone in a day. He was going to need some help. Daren was a much better metal worker than his friend Steve, but he needed his help, nonetheless. They were friends and the work was sure to go smoothly. Toby was relatively new to the trade but was a good worker and a good friend. Toby was a better metal worker than Daren.

15 The workers of the copper mine were celebrating their recent pay rises at the local pub. All of them had been awarded a bonus based on their work experience at the mine and their productivity. Darius was happy, even though he received less money than Belinda. He was still happy that he would be able to pay off a large portion of his mortgage. Lucy was also happy with her bonus. She had received more money than Darius. She was already planning the extravagant holiday she would go on.

16 Jolene was busy serving up noodles from her noodle van in the spot that she had had since the food festival began several years ago. She used to be the only one that sold noodles at the festival, but this year she had two competitors. Jolene's noodles were less expensive than Tommy's noodles. Nearby there was another noodle stand. Fanella's noodles were more expensive than Tommy's. Jolene didn't mind all that much, the festival was big enough for the three of them.

17 The town of Rockport hosted three major events that always drew in large amounts of crowds. The folk music festival was more popular than the Rockport Grand Prix. Both were important sources of income in the town. The Sailing festival was more popular than the Rockport Grand Prix, but this didn't matter to the townsfolk. Rockport was one of the most up and coming towns in the area. The yearly events were beginning to draw in crowds from further and further afield.

18 The Kingsley hills tourism board was looking to build a new museum to display the archaeological findings of the area. The board was eagerly looking for local businesses to co-fund the project. Farmer Jack's farm shop had donated less money than the Pepper-Mills hotel. The local bus company also wanted to co-fund the project, and they had donated more money than the Pepper-Mills hotel. The tourism board was sure that they would have all they needed to build the museum.

Alfred McMurphy had managed the pub ever since his father had retired several years ago. It was a family
operation and always had been. Frank McMurphy was less popular than his brother Alfred, but the patrons
liked him well enough. They were all popular when they served the drinks and listened to patrons' woes.
1 Susan McMurphy also worked at the pub and was less popular than her brother Alfred. It didn't matter of
9 course, to the McMurphy's, family always came first.

The fishermen of Rockport would stay on dry land when the regatta was taking place as no one wanted to
get in the way of the racing yachts. There was only one place they wanted to be instead, the pub. There
were many skilled captains in the local pub that day. Todd was less experienced than Herman. Herman's son
2 Melvin had recently purchased his own fishing boat and would head out with his own crew. Melvin was less
0 experienced than Todd.

There are a few teams hoping to find archaeological evidence of Dogger bank settlements that operate in the
North Sea. The Norwegian team has more funding than the British team. Their equipment is proof of this as
it's some of the most modern equipment on the market. This doesn’t mean that anyone was actively working
2 against each other though. As in all sciences, working together is key. The Danish team has less funding
1 than the Norwegian team but they still share their findings with everyone just as the British do.

The catering company that had been hired to provide food for the big pool party was having problems. Many
of their usual staff had called in sick. This meant that the company was now desperately scrambling to get
their best staff member to come in to work. Amanda was less skilled than Sammy. Both of them could use the
2 extra money. There was also another option, Jess. Jess was more skilled than Sammy. The catering
2 company boss hoped everything would run smoothly.

Three business owners from Innsmouth were looking to support the local art community and were planning to
invest in the gallery. Fred, who owned a small bistro by the beach, donated less money than Geraldine.
Geraldine was trying to be elected as mayor of Innsmouth and so had to look as though she cared about
2 local issues. Simon loved art and so was looking to pledge a decent amount of money to the gallery. He
3 donated less money than Geraldine.

It had been some time since Josephine's and Aaron's familes had come together. The two had been together
for years and now they were finally getting married. Aaron's father Gary was more popular than Aaron was.
Josephine loved to tell jokes in front of an audience. This made her more popular than Gary. The wedding
2 reception and ceremony was bound to be a battleground of sorts as to whom could give the most rousing
4 toast. This wedding would be fun for everyone.

The old chapel's doors stood wide open allowing full view of its interior. Inside there stood a magnificent,
hundred-year-old altar. There was a pulpit to the left of the altar, where the pastor would hold his sermons
from. A recent archaeological discovery has led researchers to believe that there is a secret underground
chamber somewhere to the left of the altar in the church. The local community considered the church an
2 integral part of its identity, but the number of people attending had dwindled significantly over the past few
5 years.

The mountain stood proudly as it had done for thousands of years. People had always wanted to climb it.
There was a rest stop above the camp at the mountain's base. The rest stop had over the years become
popular with bird watchers. They had no interest in reaching the summit as it was the birds they came for.
2 There was a rumour that there was a small area where people would go to throw their rubbish away below
6 the camp at the mountain's base.

The newly refurbished town surgery was finally ready to reopen after a whole month of being closed. The
surgery was located in front of a small play park. The surgery boasted more space than it had before, several
smaller buildings around were now used to house non-medical tools like cleaning equipment. The cleaner
2 would keep his floor polisher in a shed behind the surgery. The cleaner had worked at the surgery for as long
7 as anyone could remember.

2 Spice Island lay just three miles offshore. The island didn't have a lot to offer in terms of comfort but there
8 were a handful of people that called it home. There was a small weather station to the left the small dock

used to load and unload supplies. There was also a supply store to the right of the docks. It was overpriced because everything had to be imported from the mainland. Many islanders didn't mind spending more on supplies if it meant avoiding a long trip to the mainland.

The locals had long considered the caves at the bottom of the Lincoln gorge dangerous. However, the tourist industry just kept growing. An adventure holiday company had set up here too. There was a rest stop below the largest cavern. The largest cavern was usually the only place inexperienced cavers got to before turning
2 back as the tunnel that went further was a very tight squeeze. A recent geological survey of the area had
9 determined that there should be a previously undiscovered cavern below the largest cavern.

The A67 highway was the longest road in the county. Along the highway were numerous landmarks that overlooked the scenic river valley. There was a small rest stop and petrol station in front of an old bridge. The
3 rest stop was a popular place due to its great scenic overlook and saw a great many visitors. This made it a
0 profitable place to own. There was also a statue of a local folk hero in front of the rest stop.

The hall had been rented out for the high school's prom. The theme was 1950's and the hall had been decorated accordingly with bunting and posters from the era. There was a photo booth to the right of a table with snacks and refreshments. There was also a pinboard to the left of the photo booth. The pinboard was
3 where students could write inspirational quotes and messages for others to see. All the tickets for the prom
1 had sold out almost immediately.

The marathon was held every year in Dunwich. All runners ran for a charity of their choosing and would run in a group together. The runners of the local hospice were behind the runners of the wildlife charity. It didn't
3 matter who won, of course, so long as everyone had fun. The runners from the cancer charity were behind
2 the runners from the hospice. It would be a spectacular finish with a great deal of money raised for charity.

High up in the mountains, is a world-famous ski resort. The resort caters almost exclusively to the incredibly wealthy. The slope everyone at the resort desired to be seen on was above the exclusive ski lodge. People had to be specially invited to stay at the lodge. Though the lodge couldn't actually stop anyone from skiing on
3 the slope, they were responsible for its upkeep and safety. A child has gone missing below the start of the
3 slope and hadn't been seen in several hours.

The flower shop on Wensley High Street was the town's favourite. The succulents were displayed to the right of the pre-cut flowers. It was one of the reasons why the shop was so popular, people loved the fact that they could return weekly and find something new. Recently a local prankster had taken to putting fake plastic
3 plants with price tags to the left of the cut flowers, much to the dismay of anyone unfortunate enough to
4 accidentally purchase them.

The village of Northport didn't offer that much in terms of regular amusement, but it had some of the best camping around. There was a stream behind the rows of summer lodges. Children from the village would spend their entire summer playing there. They often played with those that had come to Northport for their
3 holiday. Many friendships were forged in and around that stream. There was a monument to a dog that had
5 saved his owner's life during a bear attack behind the summer lodges.

The Torton airfield was busy this time of year. The resident helicopter flight instructor was flying above the control tower. He liked to tease the people in the control tower with stunts that looked like they would end in a crash. The tower couldn't care less though. They were dealing with a small biplane that was having a few
3 engine troubles that morning. The biplane was flying below the control tower. The airfield would always get
6 busy this time of year.

The office was always busy this time of year. There was a water cooler to the left of the manager's desk. The manager had placed it there so that she could keep an eye on her underlings so that they wouldn't be tempted to chat away the whole working day. The air conditioner had broken making the office hellishly hot.
3 The manager had ordered a new fan to cool the office down. The fan was placed to the right of the water
7 cooler.

3 The little metal workshop on Yew street had been there since the 80s. There was a small desk with a chair in
8 front of a modern lathe. This place was more about metal work than paperwork, but the paperwork had to be

done nonetheless. Workers had recently noticed that some small corners of paper had been chewed off which made them suspect that a mouse was living somewhere. They believed that the mouse's hole was in front of the desk.

The mine had been a source of copper for almost 50 years. It wasn't particularly big, the company that owned it was small and couldn't afford to expand. Instead, they enjoyed a steady stream of revenue and a constant influx of highly skilled workers. Shaft-A was currently the most profitable shaft and was situated below
3 Shaft-B which was used as a storage area. There was also an exploratory drill site above Shaft-A. The
9 exploratory drilling had revealed a good copper-rich ore vein.

The food and wine festival was held every year in the Swampscott town square. It was growing year on year. There were posters for the upcoming events to the left of the entrance. Along with the usual fare of eating competitions, there were sack races, dog shows, and live music. Much to the dismay of many attendees, the
4 wine drinking competition would not be held this year. The tent used to return lost and found items to their
0 owners was to the right of the entrance.

The Rockport Grand Prix was anything but grand. It was a go-cart race for the under 15s and none of the carts could go above 30 miles per hour. The Blue team was in front of the Red team. The coaches were happy about this as they had to beat the Red and Green teams to qualify for the finals. The Green team was
4 in front of the Red team. The Grand Prix was always a popular event in Rockport and would draw in
1 onlookers from miles around.

The Kingsley hills were an area of great archaeological importance. There was a small rest stop below a fully excavated dig site. The site had yielded many interesting finds that shed light on the daily lives of early Anglo-Saxon settlers in England. There was also another potential dig site above the fully excavated one.
4 This site would soon undergo a full investigation. The archaeological work in the hills was often suspended in
2 the summer due to the influx of tourists.

McMurphy's Pub was one of the best in Amherst. The owners had decided to add additional seating so that they could serve more customers than ever before. The bar was to the right of the central row of tables. There were barstools along the whole bar and were considered to be the best seats in the house. There was
4 also an old jukebox to the left of the bar. On a Friday night, the place was full, especially if it was one of the
3 bi-weekly live music nights.

The annual sailing festival had started as it always did with the sailing regatta. Three teams from Rockport had entered the regatta in the hope of winning. The racing yacht "Royal Lady" was behind the yacht "King Fisher". The weather was perfect for sailing that day and made for much better conditions than the previous
4 years. The yacht "Fast Kitten" was behind the "Royal Lady". This was going to be an exciting race and the
4 locals cheered on the crews from Rockport as they sailed past them.

Dogger Bank in the North Sea is thought to be one of the most significant locations for evidence of early Neolithic settlements. There is only one problem, it is underwater. A team of divers and archaeologists were performing a study of the area. The submarine in charge of taking pictures was above the divers who were
4 adjusting to the pressure before they dove lower. There was also a transport submarine below the submarine
5 taking pictures. This would be used to bring anything the divers found to the surface.

The community pool was having its annual party. The water filter was to the right of the pool's ladder. This had been inspected and cleaned before the event. Staff wanted to be sure that everything was up to scratch. The pool was in better condition than ever this year due to refurbishments that had finished just before
4 summer. Though one thing was overlooked, there was a loose paving slab to the left of the pool's ladder.
6 Hopefully, no one would trip over it.

The art gallery of Innsmouth was holding an exhibition for local artists. The artist's work was on display throughout a large hall. A sculpture made of driftwood stood behind a smaller sculpture of a whale at the centre of the hall. There was also a sculpture made of washed-up plastic bottles behind the sculpture of the
4 whale. This sculpture was not very popular among the locals. Overall, the exhibition was looking to be a
7 grand success and many pieces would be sold.

The cemetery was old and within it many trees grew. On the outside wall of the chapel, there were many stone carvings. There was a gargoyle above a memorial plaque. The trees that grew in the cemetery gave plenty of shade, so it was never too hot for those that visited their dead relatives in the summer. There was
4 also a small protective symbol carved into the wall of the chapel below the memorial plaque. The trees and
8 the old stonework provided a beautiful backdrop for wedding photos.

Questions:

item	Q1-Internal	Q1-Ext	Q2
1	Who has donated neither the most, nor the least amount of money?	The Bishop sent a thank you letter to the top donor every year. Who is most likely to receive this letter?	What does the old chapel rely on?
2	Who is neither the slowest, nor fastest?	Every year there is a race to the top of the mountain, which allows only the best to compete. Who is most likely to qualify?	The time trials take place on a ________
3	Who is neither the most, nor least popular?	Patients have voted for their favourite GP. Who is likely to be the most popular?	The GP's surgery has recently______
4	Who is neither the best, nor the worst mechanic?	The least skilled mechanic is sent to pick up some coffee. Who is this likely to be?	What are the three mechanics trying to fix?
5	Who is neither the most, nor the least experienced caver?	The least experienced caver in Terrance's family is made to pack lunch. Who is this likely to be?	What does the family like to do?
6	Who is neither the most, nor the least wealthy?	A burglar wants to burgle the wealthiest person in the neighbourhood. Whose house should he choose?	What does Harriet enjoy doing every weekend?
7	Which group is neither the smallest, nor the largest?	Based on their popularity, which of the groups should have the fewest members?	What is the high school called?
8	Which of the three is neither the fastest, nor the slowest?	After the marathon, the three friends decide to compete against each other. Who is most likely to win?	The marathon draws crowds from all over the___
9	Which of the three is neither the most nor, the least wealthy?	Jonathan, Catherine, and Thom are the first to arrive at the lodge this season. Who is the least wealthy out of all of them?	What do all the people mentioned have in common?

10	Who is neither the best, nor the worst horticulturalist?	A prize is given every year to the most skilled horticulturalist. Who is most likely to win?	What kind of shop does Sam own?
11	Whose family is neither the most, nor the least respected?	A new family has moved into town and tells their son to stay away from the least respected family. Which family is this likely to be?	What season do the three friends always find a way to spend together?
12	Who has neither the most, nor the least amount of experience?	Next year Taunton airfield will host a small air show. Who is the most experienced pilot?	When did Terry start flying?
13	Who is neither the worst nor, the best cook?	The Barbeque is about to start. Which office worker is the worst cook?	What are the employees of the office organising?
14	Who is neither the worst nor, the best metal worker?	The least skilled worker in the workshop was always made to clean at the end of the day. Who is this likely to be?	How many custom jobs are the crew working on?
15	Who has received neither the biggest, nor the smallest bonus?	A tax investigation is launched that will assess workplace bonuses. Who has received the biggest bonus?	What have the workers at the mine received?
16	Who's noodles are neither the cheapest, nor the most expensive?	Young festival-goers prefer the cheapest noodles. Whose noodles are they most likely to eat?	What does Jolene's van sell?
17	Which event is neither the most, nor the least popular?	Based on their popularity, which festival is likely to have the most attendees?	What is the town called?
18	Which company had donated neither the most, nor the least?	Which local company is likely to put the least amount of money toward the tourism board's project?	What does the tourism board want to build?
19	Who is neither the most, nor the least popular?	All of the McMurphy's are well-liked by their patrons but who is the least popular among the three of them?	What is the family's last name?
20	Who is neither the least, nor the most experienced?	Who is the most skilled fisherman currently in the Rockport pub?	Why are the fishermen not fishing?
21	Who has neither the most, nor the least amount of funding?	The least funded team is going to be responsible for washing any artifacts found. Which team is this likely to be?	In what sea is Dogger bank located?
22	Who is neither the most, nor the least skilled?	If the catering company wants only the most skilled workers for the pool party who is least likely to be asked to help?	What event has the catering company been hired to cater for?
23	Who has donated neither the most, nor the least amount?	Which of the three business owners is going to donate the least amount of money to the gallery?	What is the town called?

24	Who is neither the least, nor the most popular?	Who is likely to be the least popular person at the wedding?	What event is taking place?
25	Which object is neither the furthest left, nor furthest right?	Mr. Roberts was standing in front of the altar facing it. In which direction would he go to find where the secret chamber is thought to be?	What has happened to the number of people attending service in the chapel?
26	Which location is neither at the top, nor the bottom?	A mountain ranger is at the rest stop. Which way does he need to go to find where tourists throw their rubbish?	The rest stop has become popular with which group?
27	Which location lies in the middle?	The cleaner is standing in the small play park facing away from the surgery. Which way does he need to go to pick up his floor polisher from the shed?	Roughly how long has the cleaner worked at the surgery?
28	Which location lies between the two others?	Allan is standing in front of the weather station facing it. In which direction does he need to go to get to the shop?	The island has a ____ population?
29	Which location lies between the two others?	In which general direction would Joan need to go to find the undiscovered cavern if she started at the rest stop?	What is the gorge called?
30	Which location lies between the two others?	Tom is standing by the bridge facing away from the rest stop. Which way does he need to go to get to the statue?	Roughly how long is the highway?
31	Which location is in the middle?	Which way would you have to go in order to get from the refreshments table to the pinboard?	What is the theme of the prom?
32	Which charity's team is neither first, nor last?	Which of the charity running teams is most likely to finish the marathon first?	How often is the marathon held?
33	Which location is neither at the top, nor the bottom?	The lodge staff are leaving the lodge to look for the missing child. Which way should they head?	What sort of people does the resort cater to?
34	Which one of the three items is in the middle?	Sam wants to get rid of the plastic plants, she is standing in front of the succulents. In which direction should she look to find the plastic plants?	Where is the flower shop?
35	Which location is between the two others?	A holidaymaker has lost his watch by the monument, he is standing by the stream with his back to the lodges. Which general direction should he head to find his watch?	What is the village called?
36	Who is neighter highest, nor lowest?	A photographer wants to know what is highest in the sky. What is it likely to be?	When will the airfield get less busy?
37	Which object is in between the two others?	You are standing at the water cooler facing the same direction as the manager. In which direction is the new fan?	What is the temperature in the office?

38	Which object is in between the two others?	The mouse is sitting by the lathe facing away from the desk. Which way does it need to go to get home?	Roughly when did the workshop first open?
39	Which location is in between the two others?	The foreman is in shaft B. In which direction should he take the elevator to get to the exploratory drill site?	What does the mine produce?
40	Which of the locations is in between the others?	Tommy has lost his backpack; he is standing in front of the posters facing them. Which way does he need to go to get to the lost and found tent?	Where is the festival being held?
41	Which of the three teams is neither first, nor last?	Based on the current position, which team is most likely to win?	What sort of race is the Rockport Grand Prix?
42	Which location is in between the two others?	An archaeologist has gone to the potential dig site and wants to get to the rest stop. Which way does he need to go?	For what people would the Kingsley hills be important?
43	Which object is in between the two others?	Susan is resting by the central row of tables and wants to put her favourite song on the Jukebox. Which way does she need to go?	Where is the pub located?
44	Which of the yachts is neither first, nor last?	Which sailing boat from Rockport is currently ahead of the others?	How often is the sailing festival held?
45	Which of the three is in between the other two?	The submarine used to transport finds has malfunctioned. Which way do the divers need to swim in order to fix it?	How many submarines are the archaeologists currently using?
46	Which of the three objects is in between the other two?	A lifeguard is standing in front of the water filter facing the pool. In which direction does he need to look to find the loose paving slab?	When did the refurbishments take place?
47	Which of the three pieces of art is in between two others?	Fred is standing by the sculpture made of driftwood, facing away from the whale sculpture. In which general direction does he need to go to get to the sculpture made of plastic bottles?	The gallery is currently displaying art from which sort of artist?
48	Which of the three objects is between the other two?	A spider is sitting on the gargoyle. If it wants to sit on the protective symbol, which way does it need to go?	What gives the cemetery its shade in the summer?

Chapter 6:

Chapter 6 uses the same stimuli used in Chapter 5 but with a different methodology.