

Est.
1841

YORK
ST JOHN
UNIVERSITY

Bond, Alexander Peter (2024)
Understanding the role of mentalising in reading comprehension.
Doctoral thesis, York St John University.

Downloaded from: <https://ray.yorks.ac.uk/id/eprint/12442/>

Research at York St John (RaY) is an institutional repository. It supports the principles of open access by making the research outputs of the University available in digital form. Copyright of the items stored in RaY reside with the authors and/or other copyright owners. Users may access full text items free of charge, and may download a copy for private study or non-commercial research. For further reuse terms, see licence terms governing individual outputs. [Institutional Repository Policy Statement](#)

RaY

Research at the University of York St John

For more information please contact RaY at ray@yorks.ac.uk

Understanding the role of mentalising in reading comprehension

Alexander Peter Bond

Submitted in accordance with the requirements for the degree of

Doctor of Philosophy

York St John University

School of Education, Language, and Psychology

May 2024

Intellectual Property Statement

The candidate confirms that the work submitted is his own and that appropriate credit has been given where reference has been made to the work of others.

This copy has been supplied on the understanding that it is copyright material. Any reuse must comply with the Copyright, Designs and Patents Act 1988 and any licence under which this copy is released.

© 2024 York St John University and Alexander Peter Bond

The right of Alexander Peter Bond to be identified as Author of this work has been asserted by them in accordance with the Copyright, Designs and Patents Act 1988.

Acknowledgements

Undertaking my PhD would not have been possible without the help of people at York St John University, who have been a vital area of support throughout each step of this process. There is a lovely small community in the University's psychology department, and everyone has been very easy to get hold of and helpful when I have needed advice about research, conferences or just to chat. I owe you all a great deal of thanks and I have really enjoyed the time I have spent with you all.

I would especially like to thank my PhD supervisors, Lorna Hamilton and Nicola Cutting, who have been incredibly supportive. Thank you for your knowledge, guidance, and encouragement throughout my doctoral studies. I have gained so much from you both sharing your research passions and your experiences of working in academia. The training and experience you have given me has put me in an excellent position to start an academic career. I would also like to thank my third supervisor, Jelena Mirkovic, for always being available to discuss theoretical matters and statistics (especially R). Thank you to Hannah Spring for supporting me through the process of producing a thorough systematic review, and also thanks to the rest of the psychology department, who have been incredibly motivating and encouraging.

I would especially like to thank my mother, Clare, my father, Lee, and my partner, Lisa, who have been extremely patient with me while writing this thesis. Your support has been unwavering, and for that, I am exceptionally grateful. Finally, to all of the members of the PhD community I have met over the last few years. I wish you all luck in your future endeavours!

Abstract

Mentalising is the ability to recognise that other people have mental states that influence their behaviour and actions and that these mental states may differ from one's own. Recent research suggests that mentalising may be a contributory factor in reading comprehension. Previous research has demonstrated an association between mentalising and reading comprehension in young children. However, there has been little research into the role of mentalising in adult reading comprehension, which is the focus of this thesis.

Study 1 was a proof-of-concept study that aimed to establish a link between mentalising and reading comprehension in adults. It was found that higher scores in one subcomponent of mentalising (mental state reasoning) predicted higher scores in reading comprehension, indicating that mentalising skills may be implicated in reading comprehension beyond childhood. Study 2, a systematic review, assessed mentalising tasks used in adult samples; results were used to inform task selection for Studies 3 and 4. Studies 3 and 4 investigated potential mechanisms underpinning the relationship between mentalising and reading comprehension, whereby mentalising supports the understanding of pragmatic language (indirect requests and ironic comments) during reading. To do this, Apperly's (2018) component model of perspective-taking was applied to written pragmatic language. Apperly's model proposes that when making pragmatic inferences, we take the perspective of characters at three stages. Firstly, when inferring characters' mental states (sampling); secondly, when holding this information in memory (maintenance); and thirdly, when using this information to infer the meaning of a pragmatic utterance (deployment). The results of Studies 3 and 4 indicated that mentalising is drawn upon during the first two stages of pragmatic inference (sampling and maintenance), but deployment relies on other skills (e.g., working memory). Overall, these studies indicate that mentalising supports

reading comprehension in adults, potentially with a specific role in sampling and maintaining information about characters' mental states.

List of Contents

<i>Acknowledgements</i>	<i>iii</i>
<i>Abstract</i>	<i>v</i>
<i>List of Tables</i>	<i>xiii</i>
<i>List of Figures</i>	<i>xvi</i>
<i>Abbreviations</i>	<i>xix</i>
1.1 Overview	1
1.2 General Introduction	2
1.3 What is reading comprehension?	8
1.4 Conceptualising mentalising	19
1.5 A framework for the direct involvement of mentalising in reading comprehension	27
<i>1.5.1 The relationship between narrative processing and mentalising</i>	<i>28</i>
<i>1.5.2 Approaches to studying reading comprehension</i>	<i>33</i>
<i>1.5.3 Mentalising may aid the creation of situational models (Kintsch, 1988)</i>	<i>34</i>
1.6 Chapter Summary	39
1.7 Aims and research questions	41
2. Investigating the Relationship Between Mentalising and Reading Comprehension in Adults: A Correlational Study (Study 1)	43
2.1 Chapter overview	43
2.2 Introduction	44
2.3 Method	55
<i>Participants</i>	<i>55</i>
<i>Research design</i>	<i>56</i>

<i>Materials</i>	57
<i>Procedure</i>	64
2.4 Results	66
2.5 Discussion	74
<i>Conclusion</i>	82
3. Properties of mentalising tasks used with adults: A systematic review and taxonomy	86
3.1 Overview of the chapter	86
3.2 Introduction: Development of mentalising after five years old	88
3.2.1 <i>How mentalising is measured</i>	91
3.2.2 <i>Operational terms used in this systematic review</i>	93
3.2.3 <i>The relationship between mentalising and language</i>	100
3.2.4 <i>Aims of this systematic review</i>	104
3.3 Method	106
3.3.1 <i>Eligibility criteria</i>	106
3.3.2 <i>Search strategy</i>	107
3.2.3 <i>Coding</i>	110
3.4 Results	111
3.4.1 <i>Tasks used to measure mentalising in language</i>	111
3.4.2 <i>Differences in the way mentalising tasks are described across studies</i>	112
3.4.3 <i>Operationalisations of mentalising</i>	118
3.4.4 <i>Task formats</i>	121
3.4.5 <i>Developing a taxonomy of mentalising tasks for adults</i>	123
3.4.6 <i>Associations between mentalising and language</i>	132
3.5 Discussion	145
3.5.1 <i>Measurement and conceptualisations of mentalising</i>	146

3.5.2 Using the taxonomy to examine the relationship between mentalising and language	150
2.5.3 Conclusion	152
4. When Do Comprehenders Mentalise for Pragmatic Inference? a Partial Replication Study	155
4.1 Chapter overview.....	155
4.1.1 Introduction	156
4.1.2 Pragmatic language in written text	161
4.2 Method.....	169
Participants	169
Materials.....	170
Procedure.....	176
4.3 Results.....	178
4.4 Discussion	192
Conclusions	199
5. Does Mentalising Predict Deployment of Pragmatic Inference when Interpreting Ironic Comments?	202
5.1 Chapter overview.....	202
5.1.1 Introduction	203
5.1.2 Ironic comments.....	207
5.1.3 Linguistic vs social pragmatic language	210
5.1.4 Investigating the differences between social and linguistic pragmatics	212
5.1.5 Mentalising measures and controls	216
5.1.6 Aim of the present study.....	218
5.1.7 Research questions and predictions	219
5.2 Method.....	220

<i>Participants</i>	220
<i>Materials.....</i>	221
<i>5.2.1 Norming study for the Pragmatic Language Task.....</i>	224
<i>5.1.2 Design.....</i>	235
<i>5.1.3 Procedure.....</i>	236
<i>5.1.4 Statistical analysis</i>	238
5.3 Results.....	240
<i>5.3.1 Comprehension results on the Pragmatic Language Task.....</i>	240
<i>Correlation between mentalising variables and controls</i>	243
<i>5.3.3 Analysis of predictors.....</i>	246
<i>The relationship between the Silent Films Task and whole sentence reading times in each condition.....</i>	249
5.4 Discussion	251
<i>5.4.1 Whole sentence and regional reading times: Pragmatic Language Task.....</i>	251
<i>5.4.2 Does mentalising predict reading times in the ironic comment condition?</i>	260
6. General Discussion.....	268
6.1 Summary of findings	270
<i>6.1.1 Chapter 2: Investigating the Relationship Between Mentalising and Reading Comprehension in Adults: A Correlational Study (Study 1).....</i>	270
<i>6.1.2 Chapter 3: Properties of Mentalising Tasks Used to Measure the Relationship Between Mentalising and Language in Adults. A Systematic Review and Taxonomy of Mentalising Tasks</i>	271
<i>6.1.3 Chapter 4: When Do Comprehenders Mentalise for Pragmatic Inference? A Partial Replication Study</i>	272
<i>6.1.4 Chapter 5: When Do Comprehenders Mentalise for Pragmatic Inference? Experiment 2.....</i>	274
6.2 The role of mentalising in reading comprehension in adults.....	275

6.2.1 Evidence of associations between mentalising and reading comprehension in adults	275
6.3 What do mentalising tasks actually measure?	279
6.4 Does mentalising support reading comprehension through pragmatic language?.....	282
6.4.1 Apperly's (2018) component model of perspective-taking.....	283
6.4.2 Does mentalising support pragmatic inference?	287
6.4.3 The role of working memory in deployment	290
6.5 Does mentalising have a role in the deployment of pragmatic inference?	292
6.5.1 Conclusion: Mentalising and pragmatic language	296
6.6 Limitations and considerations for future research	297
6.6.1 Future directions.....	300
6.6.2 Conclusion	301
7. Appendices	304
8. References.....	340

List of Tables

Table 2. 1 Demographic Details.....	56
Table 2. 2 Descriptive statistics for all measures and their subcomponents	66
Table 2. 3 Comparisons between SONA and Prolific participants. On measures related to reading comprehension	67
Table 2. 4 Sona - Correlations between age measures related to reading comprehension	68
Table 2. 5 Correlations between age measures related to reading comprehension.....	68
Table 2. 6 SONA and Prolific groups - Correlations Between Mentalising, Reading Comprehension, and Control Variables	69
Table 2. 7 Regression Model Predicts Variation in Reading Comprehension Using Narrative Texts.....	72
Table 2. 8 Regression Model Predicting Variation in Reading Comprehension of Expository Texts.....	73
Table 3. 1 Operationalisation of mentalising	94
Table 3. 2 The seven most frequently used tasks across 182 studies selected.....	112
Table 3. 3 Tasks examined in the taxonomy	125
Table 3. 4 Taxonomy of mentalising task properties.	128
Table 3. 5 - Reported Correlations Between the Reading the Mind in the Eyes Test and Language Abilities	134
Table 3. 6 Reported correlations between the Strange Stories task and language.....	139
Table 3. 7 Reported correlations between the Faux Pas Task and language abilities.....	143
Table 3. 8 Reported correlations between the Frith-Happé Animations Task and language	

abilities	144
Table 4. 1 Demographic statistics	170
Table 4. 2 Descriptive statistics.....	178
Table 4. 3 Correlations Between Mentalising Variables and Controls	180
Table 5. 1 Example dialogues from van Ackeren et al. (2016) stimulus set.....	205
Table 5. 2 Example narrative item and the three conditions in Marocchini and Domaneschi's (2021) study.....	213
Table 5. 3 Example item from Study 4's stimulus set, including all three conditions.....	223
Table 5. 4 The total number of responses that led to the correct interpretation of the target utterance; there were 11 participants in total	227
Table 5. 5 This table describes the total number of responses that led to the correct interpretation of the target utterance	229
Table 5. 6 This table shows the mean reading time scores for each region of the target utterance and the whole sentence reading time (Total)	241
Table 5. 7 Descriptive Statistics.....	242
Table 5. 8 Mean Scores for mentalising variables and controls.....	242
Table 5. 9 Correlation Between Mentalising Variables and Control Variables.....	243
Table 5. 10 Correlations Between Mentalising and Control Variables and the Reading Times on the Pragmatic Language Task	244

List of Figures

Figure 1. 1 Direct and Indirect Effects Model of Text Comprehension (Kim, 2016; 2017) ..	14
Figure 1. 2 Kintsch (1988) Construction Integration Model	16
Figure 1. 3 The Direct and Indirect Effect Model of Reading (DIER; Kim, 2017).....	18
Figure 1. 4 The Unexpected Transfer Task by Baron-Cohen et al. (1985).....	21
Figure 1. 5 Referential Communication Task (the Director’s Task)	25
Figure 1. 6 Alignment between mentalising development and children’s narrative processing abilities	29
Figure 2. 1 The display screen in the Director’s Task	58
Figure 2. 2 The Reading the Mind in the Eyes Test stimuli	59
Figure 2. 3 A vignette from the Strange Stories	60
Figure 2. 4 Example passage from the Implicit Management Task.....	64
Figure 3. 1 Flowchart of the process of limiting publications from the initial database search	109
Figure 3. 2 Operationalisations of mentalising across the 182 studies	119
Figure 3. 3 The distribution of different tasks used to measure mentalising in the studies reviewed.....	121
Figure 3. 4 False belief (mentalising) cartoons.....	122
Figure 4. 1 Apperly’s (2018) Component Model of Perspective Taking	158

Figure 4. 2 Aware and Unaware Conditions in the Inference Group from Trott and Bergen’s (2020) stimuli	163
Figure 4. 3 Trott and Bergen’s (2020) mixed design study	165
Figure 4. 4 Trott and Bergen’s (2020) Pragmatic Language Task (Study 2).....	171
Figure 4. 5 Figures show the checks for homogeneity of variance, multicollinearity, influential observations (outliers), and normality of residuals	179
Figure 4. 6 Comparison of the results obtained in this study with those found in Trott and Bergen’s (2020) study.....	182
Figure 4. 7 Interaction between the awareness condition and the Knowledge and Inference Groups.....	184
Figure 4. 8 Density charts of individual differences in mentalising ability across both experimental groups.....	187
Figure 4. 9 Interaction between awareness and working memory in the Inference Group ..	191
Figure 5. 1 Example screenshots of the indirect request condition in Marocchini and Domaneschi’s (2022) study	214
Figure 5. 2 Presentation sequence for each item in the norming study	226
Figure 5. 3 Example screenshots of the indirect condition in Study 4.....	238
Figure 5. 4 The relationship between the Strange Stories and whole sentence reading times in each condition	247
Figure 5. 5 The relationship between the Silent Films Task and whole sentence reading times in each condition.....	249
Figure 6. 1 The Dot Perspective Task (Samson et al., 2010).....	286

Abbreviations

ADHD	Attention deficit hyperactivity disorder
ART-2	Adult Reading Test, 2 nd Edition
DIER	The direct and indirect effects model of reading
DIET	Direct and Indirect Effects Model of Text Comprehension
ERIC	Institute of Education Science
ESCoT	Edinburgh Social Cognition Test
fMRI	functional Magnetic Resonance Imaging
GCSE	General Certificate of Secondary Education
ICCs	Interclass correlations
IMCs	Instructional manipulation checks
LexTALE	The Lexical Test for Advanced Learners of English
LMMs	Linear mixed models
PsyArticles	American Psychological Association
RMET	Reading the Mind in the Eyes Test
SPIDER	Sample, phenomenon of interest, design, evaluation, research type
TOM	Theory of mind

Chapter 1. Introduction

1.1 Overview

Fictional novels can have a profound effect on people. When reading about fictional worlds, people often form strong emotional bonds with the characters. By experiencing how characters react to moral dilemmas and adversity, readers can change their attitudes about the world and even reshape their beliefs about themselves (Appel & Mara, 2013; Escalas, 2004; Djikic et al., 2009; Richter et al., 2014). However, although it is widely accepted that reading gives us the opportunity to learn about the experiences of others, it is easy to underestimate how much of our own experience we bring to the task of reading. Novels do not describe every thought each character has at every moment; instead, readers are left to fill in the blanks. This allows the reader to create fully-fledged characters from only brief descriptions of the character's behaviour. In other words, the reader uses information from the text and combines it with their own experiences and knowledge of the way people think, feel, and behave to create rich

mental representations of characters. Because the reader's unique personal experience is incorporated into the characters during reading, two people reading the same book, such as *Pride and Prejudice* (Austin, 1813), may have very different ideas about what Mr. Darcy is really like. As the plots of novels are often driven by the internal thoughts and motivations of the characters, people who are more skilled at reasoning about thoughts and feelings may have a greater understanding of the characters and the novel as a whole.

This thesis examines how mentalising, as a form of social cognition, contributes to individual differences in adult reading comprehension.

1.2 General Introduction

Reading comprehension is crucial for effective participation in education, employment, cultural and social activities (Tobia & Bonifacci, 2015). Yet a 2015 survey in England revealed that one in six adults leaves school with very poor reading comprehension levels, which means they may struggle to read longer texts and/or texts on unfamiliar topics (OECD, 2019). Recent surveys suggest that Adults with poor reading face greater struggles in the workplace. In the United Kingdom, a 2011 Confederation of British Industry survey of 566 employers highlighted that 42% of respondents were unsatisfied with workers' basic use of English, and 44% of employers had to invest in remedial training for school and college leavers (CBI, 2011). As a link between literacy levels and economic outcomes has been found in adulthood, it has been suggested that re-training opportunities must be made available to adults with poor literacy skills in order to give them more opportunities in the workplace (Evans & Hares,

2021). As reading comprehension is a vital skill for adults' quality of life, it is important that the factors that contribute to poor reading skills are understood.

Reading comprehension can be defined as the process of extracting and constructing meaning from written texts (RAND Reading Study Group 2002, p. xiii). It is a multifaceted skill that is supported by a number of other sub-skills (e.g., vocabulary knowledge, grammar, inference generation); therefore, if children struggle with one of the skills that support reading, their overall reading ability may be affected as a result (Kendeou et al., 2014; Oakhill & Cain, 2012; Royeras & Sumayo, 2024). Research suggests that adult reading problems usually stem from individuals falling behind in early childhood (Hernandez, 2011; Oakhill & Cain, 2012). However, these skills can still be improved in adulthood to improve reading comprehension across childhood and adulthood. For example, reading fluency is the ability to read text with accuracy, speed, with accurate pronunciation, and has been shown to contribute to reading comprehension (Metsala & David, 2021). Eloranta et al. (2018) found that adults diagnosed with reading disabilities in childhood have been shown to improve their reading to a level where criterion where adult age reading disabilities were no longer met. As reading comprehension draws on many sub-skills, the source of comprehension difficulties may vary and require targeted interventions. As such, it is crucial to understand the different cognitive skills that contribute to individual differences in reading comprehension.

The Simple View of Reading (SVR) is an influential model, which describes the sub-skills that underlie reading comprehension. Specifically, the SVR conceptualises reading comprehension as the product of word decoding (rapidly matching letter[s] to corresponding sounds) and oral language comprehension (ability to derive meaning from spoken discourse; Catts et al., 2006; Hoover & Gough, 1990; Ouellette, 2006). Studies suggest that both skills have a substantial impact on reading comprehension in young

children (Kendeou et al., 2009; Hoover, 2023). However, after children become fluent word readers, which is normally achieved by 10 years old, the importance of oral language comprehension to reading comprehension increases (García & Cain, 2014; Language & Consortium, 2015; Lonigan et al., 2018). Oral language comprehension comprises a complex subset of skills ranging from word-level (e.g. vocabulary knowledge) to higher-level discourse processes. Recently, attention has turned to how social cognition, and specifically mentalising, might contribute to oral language comprehension and, in turn, reading comprehension (Kim, 2017). As the importance of oral language ability in skilled reading comprehension increases with age, it is possible that mentalising may be an important driver of individual differences in adult reading (Kim et al., 2016, 2017; Massonnié et al., 2017).

Mentalising may be particularly important in the comprehension of fictional texts, i.e., narratives that are often propelled by the mental states and motivations of the characters. When reading fictional texts, mentalising may be drawn upon to represent and monitor information about characters' mental states; this allows the reader to make inferences about the characters' motivations and behaviours even when mental state information is not explicitly mentioned, thereby improving their understanding of the text (Dore et al., 2018). Indeed, there is evidence that by the age of four, children track the spatial and mental perspectives of characters while listening to stories (Fecica & O'Neill, 2010; O'Neil & Shultis, 2007; Rall & Harris, 2000; Ziegler et al., 2005). The ability to track characters' perspectives may vary from person to person, which is likely due to individual differences in mentalising ability (Cartwright, 2015). Readers with better mentalising abilities may have a better understanding of the character's mental states and thus the motivations for the character's actions in the narrative as well.

Up to this point, only a small pool of research has investigated the link between mentalising and reading comprehension, and this research has focused on young children (Atkinson et al., 2017; Ebert, 2020; Kim, 2016, 2017). It is possible that mentalising has an important role in reading comprehension in young childhood but less so in adulthood. During young childhood (two to five years old), children develop an understanding of several important mentalising concepts, such as the knowledge that different people can hold different desires and beliefs (see Section 1.4 for more detail) (Wellman & Liu, 2004). The most crucial stage of conceptual understanding at this age (usually gained from four to six years old) is a knowledge of false beliefs, which means understanding that other people can hold subjective and faulty representations of the world (Apperly, 2012). An understanding of false beliefs may be particularly important when inferring characters' motivations in narrative texts. For example, in the fairy tale Hansel and Gretel, the two children are left hungry and tired in the woods. A witch invites them into her gingerbread house and feeds them. Hansel and Gretel have the false belief that the witch has benevolent intentions, but in fact, the witch intends to trap them and eat them. A child with an understanding of false beliefs may better understand how Hansel and Gretel's false beliefs about the witch's intentions drove them to enter the witch's gingerbread house (Grimm & Grimm, 1812). In support of this assertion, recent longitudinal studies have shown that children with a better understanding of false beliefs at four years old predicted better reading comprehension ability at age six (Atkinson et al., 2017). Ebert (2020) also showed that better false beliefs understanding was associated concurrently with language skills (receptive grammar, sentence comprehension) that support reading comprehension in five-year-olds. However, false belief understanding at five years old did not predict reading comprehension longitudinally at age 12. In both Atkinson and Ebert's studies, the same two false belief

tasks were used, and this may suggest that understanding false beliefs earlier may give children an advantage when reading at age six. On the other hand, at age 12, when children have had more reading comprehension practice and tuition, the advantages of understanding false beliefs at an early age are no longer seen. Therefore, it may be that early understanding of mentalising concepts (e.g., false beliefs) predicts better reading comprehension, but after children have grasped these concepts, the effect of mentalising on reading comprehension may not be found.

Alternatively, mentalising may continue to have a role in reading comprehension in adulthood. Although individuals may gain *knowledge* of crucial mentalising concepts in childhood (e.g., false beliefs), research indicates that individuals' ability to *use* this conceptual knowledge effectively to understand the mental states of others appears to continue to show variability in adulthood (Apperly, 2012) (see Section 1.4 for more detail). Therefore, it may be that adults who are better at using their knowledge of mental state concepts are better at understanding the thoughts, feelings, and motivations of characters in literary texts (Bruner, 1986). Even in adulthood, processing the mental states of other people who have different perspectives seems to be more cognitively taxing than processing perspectives that align with one's own. For example, studies have shown that adults are quicker at processing the beliefs of others when they hold a true belief about an object (the red ball is red) than when another person holds a false belief about an object (the red ball is yellow) (Apperly et al., 2008). This would suggest that even though adults understand that others can hold false beliefs, they may find it hard to use this knowledge to interpret the beliefs of others when they differ from their own understanding of reality. Furthermore, there is evidence that the ability to reason about the mental states of others appears to vary in adulthood. For example, variability can be seen in adults' ability to interpret false beliefs, to reason about mental states, and their

ability to make attributions from non-verbal cues (Fairchild & Papafragou, 2021; Rahman et al., 2021). Therefore, variability in adults' ability to use mental state concepts to understand the motivations of characters may explain some individual variance in reading comprehension in adults.

More research is needed to establish whether a relationship between mentalising and reading comprehension exists in adults and, if so, what the mechanisms of this relationship are. Given that mentalising is thought to support reading comprehension via oral language comprehension and the importance of oral language comprehension to reading is thought to increase as we age, it is possible that mentalising ability is particularly important for reading comprehension in adults (García & Cain, 2014; Language & Consortium, 2015; Lonigan et al., 2018). Hence, this thesis aims to explore the role of mentalising in typically developing adults.

This chapter begins by describing cognitive and developmental accounts of reading comprehension and mentalising, followed by consideration of theoretical accounts of the relationship between mentalising and reading comprehension. Furthermore, empirical research that has indicated a relationship between mentalising and reading comprehension is evaluated. Chapter 2 reports an empirical study investigating whether a relationship between mentalising and reading comprehension exists in adults using three different mentalising measures. This “proof-of-concept” study in Chapter 2 was necessary to address the gap in the literature as a relationship between mentalising and reading comprehension has, so far, only been established in children, but not adults. However, because only one of the three mentalising measures predicted unique variance in reading comprehension, a systematised review was conducted to better understand the different mentalising measures that have been used in research with adults. In the systematic review in Chapter 3, tasks used to measure

mentalising in adults are collated and examined in a taxonomy to provide a better understanding of each task's properties and which sub-components of mentalising they capture. Chapter 4 and Chapter 5 report two experimental studies, aiming to explore potential mechanisms underpinning the relationship between mentalising and reading comprehension. Finally, Chapter 6 discusses the overall findings of the thesis.

1.3 What is reading comprehension?

Reading comprehension has been defined as 'the processes of simultaneously extracting and constructing meaning through interaction and involvement with written language' (RAND Reading Study Group 2002, p.xiii). In the SVR the two components (word decoding and oral language) are weighted equally and recognised as independent abilities supported by different sub-skills. Recent research indicates that the SVR accounts for a large proportion of the variance in reading comprehension both concurrently and longitudinally (Hjetland et al., 2018; Kim, 2017). In a large concurrent study in the United States, Kim (2017) determined that 100% of the variance in reading comprehension in English among young children was explained by oral language and word decoding ($M \text{ age} = 7.54, SD = .64$). Furthermore, there is also causal evidence that the SVR predicts reading comprehension ability longitudinally. Studies have suggested that word decoding and oral language comprehension predict reading comprehension from four to six years old, from six to 10 years old, and from 10 to 13 years old (Juel, 1988; Kendou et al., 2009; van Wingerden et al., 2018). The SVR is also generalisable to other languages, such as Dutch, Italian, Norwegian, and Chinese (de Jong & van der Leij, 2002; Hjetland et al., 2018; Ho et al., 2012; Tobia & Bonifacci, 2015). For example, research in Norwegian has supported that oral language and word decoding

skills at age four predicted reading comprehension at age seven longitudinally, with concurrent indices of language and decoding explaining 99.7% of the variance in reading comprehension at age seven (Hjetland et al., 2018). Therefore, both concurrent and longitudinal studies suggest that word decoding and oral language comprehension are the sole factors involved when accounting for reading comprehension skills.

While evidence suggests that the SVR is effective at predicting reading comprehension ability, there are some studies that question the validity and generalisability of this theory. Indeed, a substantial amount of research suggests that variance in reading comprehension cannot be explained by the SVR variables alone. Overall, studies using SVR indicators explain between 41% and 100%, of the variance in reading comprehension (Lonigan et al., 2018; Kim, 2017). Some of this variation may be due to researchers using different types of tasks to measure reading and language skills. For example, word decoding has been measured using non-word tasks and (real) word recognition tasks, but non-word reading—a test of phonological awareness—may be more suitable for beginner readers, while word recognition is better for more proficient, older readers whose phonological skills are more developed (Conners, 2009; Hoover & Gough, 1990). This may explain why some studies using non-word reading as a decoding measure have concluded that the SVR explains only 48% of the variance in reading comprehension among children 8–10 years old; instead, a word recognition task may be more appropriate for this age group (Georgiou et al., 2009). Joshi and Aaron (2000) demonstrated that word naming speed predicts reading comprehension above and beyond word decoding and oral language comprehension, but these results have only been replicated when using non-word reading tasks as a decoding measure (Cutting & Scarborough, 2006; Johnston & Kirby, 2006). Therefore, due to the methodological issues in Joshi and Aaron's study, it cannot be concluded that naming speed should be

included as an additional variable in the Simple View of Reading. Nevertheless, their results suggest that variables beyond word decoding and oral language may explain additional variance in the model. Overall, these findings point to the possibility that at least some of the unexplained variance in the Simple View of Reading can be explained by differences in the types of tasks used to measure the component processes.

To understand the variance in reading comprehension left unexplained by the SVR, researchers have suggested that different third variables that directly contribute to reading comprehension may be missing from the model (Atkinson et al., 2017; Conners, 2009; Ricketts et al., 2013). Conners (2009) posits that attentional control should be added as a third variable to account for metacognitive strategies, as attention to one's own comprehension of the text is needed to implement compensatory reading strategies. For example, by the age of eight, children can form metacognitive strategies, such as pausing and re-reading, to assist their comprehension when their understanding of the text is inefficient (Walczyk et al., 2004). Thus far, studies have not been able to replicate Conners' results, which implies that attention does not make a unique contribution to reading comprehension (Borella & De Ribaupierre, 2014; Christopher et al., 2012). However, other studies have suggested that attention supports working memory processes during reading comprehension, in part through attentional control of intrusive thoughts (McVay & Kane, 2012). Kim (2016) proposes that attention, via working memory, may contribute to the oral language component of the SVR.

Social cognitive skills such as mentalising may help readers adopt strategies to improve their comprehension of the text. Savage (2006) examined the relationship between the components of the SVR and reading comprehension in 56 adolescents (M age = 15) with poor reading comprehension. Savage concluded that these adolescents had poor word-reading ability, which they compensated for by using context to

understand unfamiliar words or phrases. It is possible that readers use social context (e.g., speaker's intentions, motivations) in a similar way to understand ambiguous phrases. Social context may enable readers to understand the intentions behind non-literal spoken utterances, such as indirect requests. Accordingly, it has been suggested that mentalising may be drawn upon (along with language skills) when reading spoken dialogue that draws on mental states (e.g., knowledge, thoughts, emotions; Andrés-Roqueta & Katsos, 2020). Therefore, mentalising may be a potential additional component variable in the SVR (discussed further in Section 1.4).

Atkinson et al. (2017) present longitudinal evidence indicating a potential causal relationship between mentalising and reading comprehension. They found that false belief understanding at age four predicted performance on standardised fiction and non-fiction reading comprehension tasks at age six when controlling for age, executive functions, and the two components of the SVR (i.e., oral language and word decoding). Atkinson et al. (2017) argue that early acquisition of false belief understanding may facilitate the development of other metacognitive abilities linked to reading comprehension, such as metamemory (self-monitoring and self-control over memory processes), source monitoring, and comprehension monitoring (Lecce et al., 2014; Lecce & Bianco, 2019; Lockl & Schneider, 2007). As such, gaining an early understanding of false beliefs may give children more opportunities to use metacognitive strategies that aid their higher-order (oral or reading) comprehension skills.

As Atkinson et al. (2017) found that mentalising predicted reading comprehension above and beyond the components of the SVR, they argue that mentalising should be added as a component within the SVR. In support of this claim, Ricketts et al. (2013) demonstrated that advanced mentalising measures (Strange Stories and Frith-Happé animations) also predicted concurrent performance in reading

comprehension above and beyond oral language comprehension and word decoding in autistic adolescents. Taken together, these studies indicate that mentalising may support reading comprehension across development from young childhood to adolescence. Atkinson et al. (2017) maintain mentalising and meta-cognitive processes (e.g., comprehension monitoring) may be involved in building situational models by aiding the integration of textual information during reading (Kintsch, 1988; Prinz et al., 2020). Overall, these studies provide evidence of a special relationship between mentalising and reading comprehension, although more evidence is needed before any explicit conclusions can be made about the mechanisms that underpin this relationship (Atkinson et al., 2017; Kim, 2017; Ricketts et al., 2013).

A criticism of the SVR is that, although word decoding and oral language comprehension are equally weighted within the model, the importance of each component appears to vary with age (Catts et al. 2005). García and Cain's (2014) meta-analysis of 110 studies revealed that the power of word decoding strongly predicts reading comprehension until age 10, whereas after this age, there is a reduction in the strength of correlation. The opposite relationship emerges for oral comprehension skills and appears to explain greater variance in reading comprehension as age increases (Language & Consortium, 2015). This pattern of results occurs because when children first start to read, understanding the relationship of letters to sounds (word decoding) is the most challenging aspect of reading; however, as this ability becomes more automated, individual differences in oral language account for greater variance in reading comprehension (Adlof et al., 2005).

Longitudinal research has shown that the variance of reading comprehension accounted for by oral language increases with age. A longitudinal study conducted in the United States by Catts et al. (2005) tested 527 poor readers at ages seven, nine, and 13

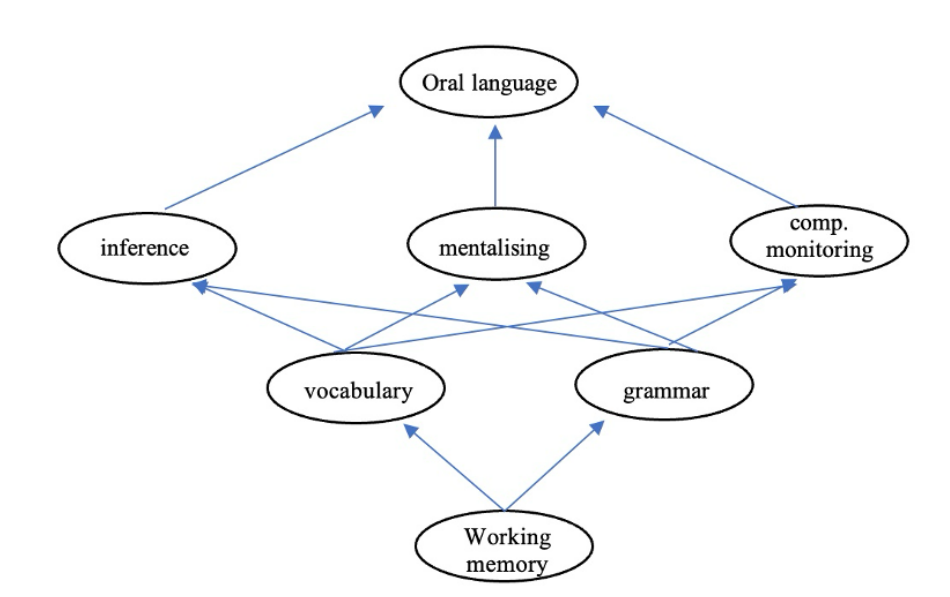
and found that the components of the SVR accounted for approximately 75% of the variance in reading comprehension during this period. Further analysis indicated that word decoding and oral language accounted for both shared variance (approximately 40%) and unique variance (approximately 35%) in reading comprehension. However, the amount of unique variance accounted for by each of the components differed across this developmental period. Specifically, the unique variance in reading comprehension explained by word decoding dropped from 27% at age seven to 13% at age nine to just 2% at age 13. In contrast, the amount of unique variance explained by oral language comprehension increased from 9% to 21% and then to 36% during this same period. Catts et al. (2005) posit that the skills that contribute to oral language are much less understood than those that support word decoding, and most oral language interventions consist of vocabulary training, which is insufficient. Therefore, a more accurate conceptualisation of oral comprehension skills may provide a more holistic understanding of the processes involved in reading comprehension (Massonnié et al., 2018).

Kim's (2016, 2017) work has attempted to develop a better understanding of the processes and structural relations in the SVR by unpacking the skills involved in oral language comprehension. Although the SVR explains a large proportion of the variance in reading comprehension, the original authors admit that the two key components (word decoding and oral language comprehension) of the model comprise complex processes in and of themselves (Gough et al., 1996). Kim (2016, 2017) has expanded upon the sub-skills underpinning the oral language branch. Furthermore, the SVR describes the structure of components involved in reading comprehension but does not explain how the two components (or the skills they comprise) interact during the process of reading comprehension. Hence, in addition to describing the skills involved in oral language

comprehension, Kim describes how these skills interact during reading. Kim's work, therefore, provides a fuller understanding of the processes and structures involved in oral language and reading comprehension.

Figure 1. 1

The Direct and Indirect Effects Model of Text Comprehension (Kim, 2016, 2017)



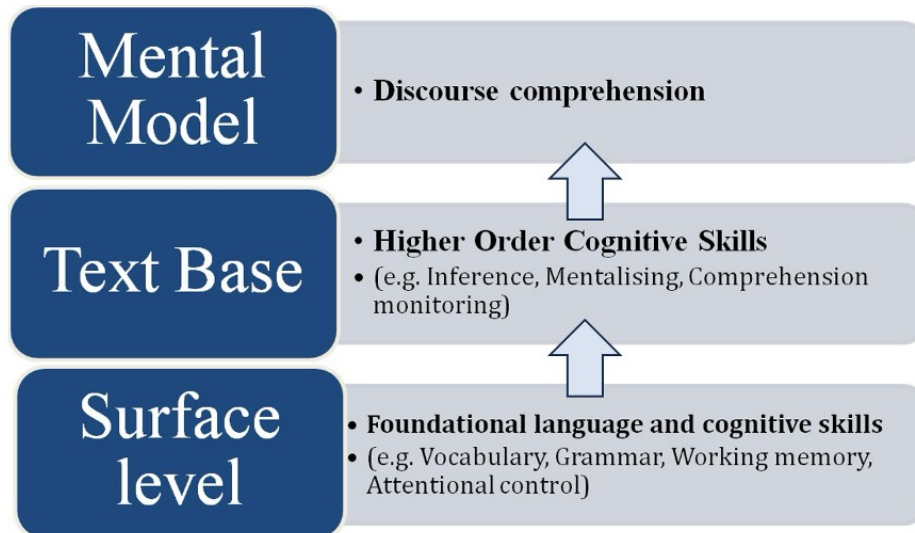
Note: In this model, lower-level processes supported by working memory, vocabulary, and grammar feed into higher-level processes supported by inferencing, mentalising, and comprehension monitoring.

To understand the relationships between the component skills of oral language, Kim developed the direct and indirect effects model of text comprehension (DIET; 2016, 2017), which proposes that the oral language factor in the SVR is composed of a hierarchical set of cognitive and language skills (see Figure 1.1). Kim (2016) maintains that interpreting oral language is a dynamic process where meaning is gained at the lexical, sentence, and discourse levels. Therefore, several different low- and high-order skills are needed to interpret meaning at each level.

This dynamic view of oral language comprehension is based on Kintsch's construction integration model (1988) of discourse comprehension. In this model, there are three levels of comprehension. (1) At the surface level, the reader decodes and interprets the meaning of individual words. (2) At the text base level, the meaning of sentences is comprehended, and links are made across sentences. Finally, (3) the situational level is the highest level of comprehension, where the literal meaning of the text is integrated with the reader's background knowledge to form a mental representation of the text. In the DIET model, Kim attempts to describe the skills needed at each level (surface, text base, situational) to achieve oral language comprehension (See Figure 1.2). Low-level processes (working memory, vocabulary, grammar) feed into higher-level processes (inferencing, mentalising, comprehension monitoring), which support oral language comprehension.

Figure 1. 2

Kintsch's (1988) Construction Integration Model



Note. The Construction Integration Model was originally designed by Kintsch (1988). Kim (2016, 2017) used the Construction Integration Model as a conceptual model for her Direct and Indirect Effects Model of Text Comprehension (DIET). In the DIET model, language and cognitive skills are mapped onto the three levels of mental representations: surface code, text base, and situation model.

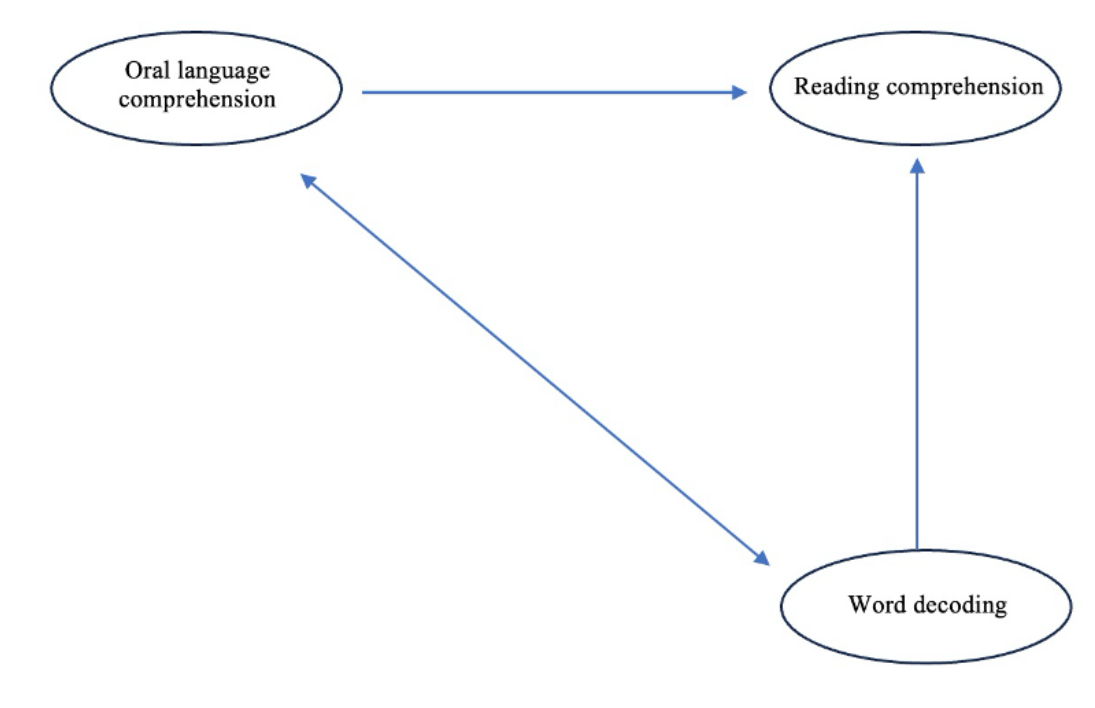
Kim (2017) reported that component sub-skills of the DIET model explained 86% of the variance in children's language comprehension. In this cross-sectional study, 350 English-speaking children's (M age = 7.5) foundational cognitive skills (working memory), foundational language skills (vocabulary, grammar), and higher-order cognitive skills (comprehension monitoring, inferencing, mentalising) were measured. Direct relationships were found between the foundational cognitive skills, foundational linguistic skills, and higher-order cognitive skills, suggesting that the lower levels of comprehension feed into the higher levels. Notably, mentalising abilities (second order

false beliefs) and comprehension monitoring were directly related to oral language comprehension, whereas inferencing was only indirectly related to oral language comprehension via mentalising and comprehension monitoring. This indicates that the ability to reason about the intentions and perspectives of others (mentalising) is particularly important to understanding oral texts, possibly more so than the ability to combine oral language information with general background knowledge (inferencing). Furthermore, Kim (2016) demonstrated that the DIET model explains 85% of the variance in oral language in Korean, suggesting that the DIET model (like the SVR) can be generalised cross-culturally.

The direct and indirect effects model of reading (DIER; Kim, 2017) was developed to understand how the oral language component of the SVR (explored in the DIET model) combines with word decoding to support reading comprehension. To explore the pathway structure of the DIER model, Kim (2017) analysed the indirect and direct relationships between the skills within the model in 350 English-speaking young children.

Figure 1. 3

The Direct and Indirect Effects Model of Reading (Kim, 2017)



The subcomponents of the DIER model (see Figure 1.3) explained 100% of the variance in reading comprehension. Oral language comprehension was moderately associated with reading comprehension, whereas word decoding was strongly associated with reading comprehension in seven-year-old children. These results reflect previous research supporting that word decoding has a predominant position within the SVR until the age of 10. However, after this age, the influence of oral language increases (García, & Cain, 2014). Overall, the results suggest that reading comprehension relies on foundational cognitive and language skills that feed into higher cognitive abilities, and both these foundational skills and higher cognitive abilities are indirectly related to reading comprehension via oral language comprehension and word decoding. Notably, mentalising was the only higher-order cognitive ability that was directly associated with

reading comprehension, suggesting it may have a key role in understanding the mental states of characters.

In summary, Kim's (2016, 2017) work has expanded the SVR by describing the component skills within the oral language branch, illustrating how they interact together to support reading comprehension. Importantly, Kim's work indicates that mentalising may have a prominent role in supporting oral language comprehension and, consequently, reading comprehension. However, most research to date has investigated this relationship in children under the age of 10 in the process of learning to read. Given that the role of oral language comprehension in the SVR increases in importance with age and the relationship between oral language and mentalising, it is reasonable to infer that the role of mentalising in reading comprehension may also increase in importance with age. The same relationship can be seen in the development of another component of oral language, vocabulary knowledge, which also continues to grow over one's lifespan through increased exposure to language (Kavé et al., 2020). Extending this research to adult populations can, therefore, expand the current understanding of the relationship between mentalising and reading comprehension as well as how this relationship may change over time.

1.4 Conceptualising mentalising

Mentalising is the ability to recognise and reason about the mental states (e.g., thoughts and feelings) of oneself and others that underpin behaviour (Wellman et al., 2001). It is often used synonymously with other terms, such as the theory of mind and mind-reading (Rosenbaum et al., 2015). Mentalising allows individuals to reason about

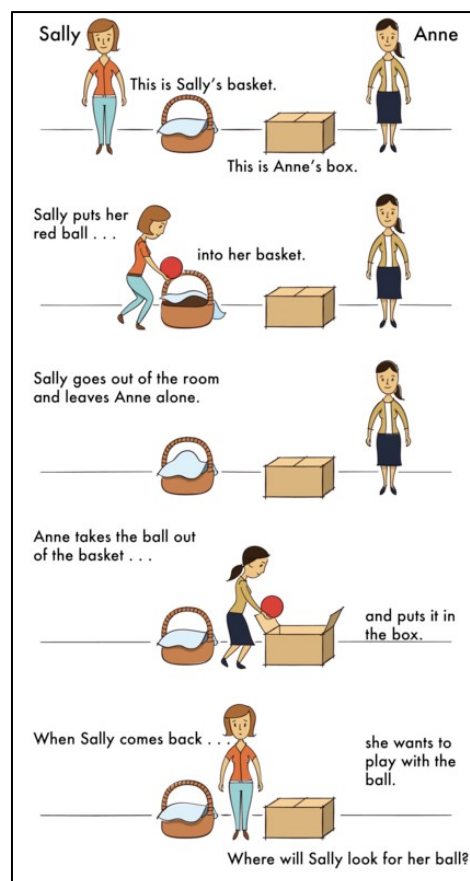
the intentions underlying people's actions and language. Using this information, people can predict, explain, and influence other people's behaviour, which is vital when working collaboratively to achieve shared goals (Heyes & Frith, 2014). The development of mentalising has traditionally been described in terms of conceptual stages. Researchers have described a universal pattern of development in young children up to five years defined milestones of conceptual understanding (e.g., understanding false beliefs) that are reached by this age (Astington & Jenkins, 1999; Liu et al., 2008; Rakoczy, 2022; Wellman & Liu, 2004). From the ages of two to five years old, children develop an understanding first that different people can hold different desires, then that people hold different beliefs, and after this, that some people can be knowledgeable or ignorant about a given topic; finally, children gain an understanding of false beliefs (Wellman & Liu, 2004).

A key focus of mentalising research has been the age at which children gain an understanding of false beliefs or that people can hold subjective and inaccurate representations of reality (Apperly, 2012). False belief understanding is usually gained at around four years old (Rakoczy, 2022). Using this knowledge, researchers have been able to associate mentalising development in young children with the development of other skills by correlating the age at which children grasp an understanding of false beliefs with other domains of development, such as language (Slade & Ruffman, 2005; Wellman, Cross, & Watson, 2001). Understanding false belief is an important developmental shift as it enables children to understand that others may have different perspectives from their own and permits them to represent (imagine) people's attitudes toward other people or objects (Perner, 1991). This is an especially useful skill when working cooperatively with other children or understanding the perspectives of characters in narratives (Johnson-Laird, 1983; Perner, 1991).

False belief tasks are often administered as a story, which can be read to participants or read by participants. These tasks are measured using a simple pass/fail scoring system, indicating that the child either does or does not have an understanding of false beliefs. An example of a false belief task is the unexpected transfer task (see Figure 1.4), which tests whether children understand that Sally can have a false belief about the location of a ball.

Figure 1. 4

The Unexpected Transfer Task by Baron-Cohen et al. (1985)



Alternative false belief tasks try to minimise how much the tasks draw on language. For example, in an unexpected contents task, children have to guess the contents of a

Smarties container (to which they usually say ‘Smarties’) (Hogrefe et al., 1986). After this, they are shown that the tube contains coloured pens. They are then asked to guess what their friend would say if they were asked to name the contents of the tube. If the child says ‘Smarties’, it shows they understand that other people can have false subjective views of the world. False belief tasks show good reliability and validity, and regardless of the type of false belief task (e.g., unexpected transfer task or unexpected contents), performance is coherent, meaning that performance on one false belief task reliably predicts performance on another (Warnell & Redcay, 2019; Wellman et al., 2001).

After children have developed false belief understanding at around five years, there are few clear signs of age-related conceptual milestones of mentalising development. This may be because conceptual *knowledge* development is more continuous, and the ability to *use* mentalising knowledge to interpret social situations is increasingly supported by other cognitive skills (e.g., language and memory; Apperly, 2012). For example, it has been shown that after children develop an understanding of false beliefs, they then continue to develop an understanding of second-order false beliefs (understanding that Person A can hold a false belief about Person B’s belief; Astington et al., 2002; Rakoczy, 2022). Astington et al.’s (2002) longitudinal study determined that from the age of five to seven years old, there is a significant improvement in children’s second-order false belief reasoning, above and beyond abstract reasoning and general language. These findings suggest that conceptual knowledge development occurs past the age of five years old. However, it is likely that other cognitive skills support an understanding of second-order false beliefs. Holding false beliefs in hierarchical structures draws on working memory capacity. Comprehending second-order false belief sentences (e.g., *Desmond thinks that Molly*

believes that it is still raining) may draw on increasingly sophisticated syntax knowledge. Indeed, intervention studies have supported that working memory and syntax training improve second-order false belief understanding from five to seven years old (Polyanskaya et al., 2022). Hence, although there may be a development in conceptual false belief knowledge from five to seven years old, developments in other cognitive skills (language, memory) also support the understanding and use of second-order false beliefs.

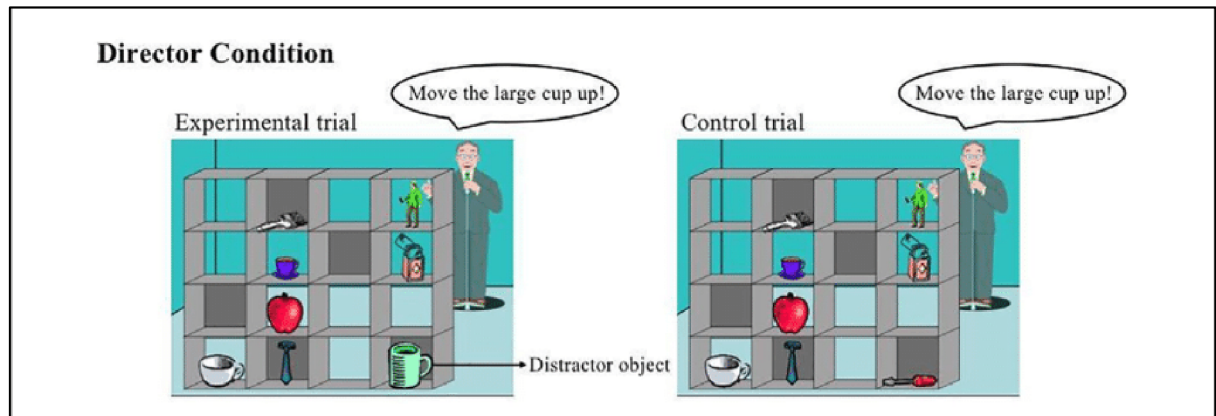
After the age of seven, researchers have examined mentalising ability through more indirect means, using ‘advanced’ mentalising measures that test individuals’ understanding of non-literal language or their ability to interpret the intentions behind people’s actions (Devine & Hughes, 2013, 2016; White et al., 2009). These more complex measures use vignettes or short film clips depicting social scenes followed by questions, usually asking why someone said something or acted in a certain way. Advanced mentalising tasks capture individual differences, and performance increases with age in typically developing children (Devine & Hughes, 2015; Happé et al., 1998; Hughes, 2011). For example, O’Hare et al. (2009) found that children can only competently reason about misunderstandings using their knowledge of second-order beliefs at the age of nine years old and can understand the intentions behind non-literal speech such as persuasion and sarcasm by age 12. Indeed, individual differences in advanced mentalising are apparent across the lifespan, indicating that individuals do not fully master the use of mentalising concepts even in adulthood (Cavallini et al., 2013; Marocchini & Domanseschi, 2022; Wang & Su, 2006). Notably, advanced mentalising tests do not represent milestones of conceptual development. They merely measure an individuals’ ability to use their knowledge of standard mentalising concepts (e.g., understanding desires, beliefs, knowledge vs ignorance, and false beliefs) in a more

sophisticated manner (Apperly, 2012). Indeed, advanced mentalising tasks have shown increasing proficiency in the appropriate *use* of mentalising abilities across middle childhood (Apperly, 2012; Devine & Hughes, 2016). Therefore, Apperly (2012) contends that advanced mentalising tasks test individual differences in harnessing cognitive skills (e.g., language skills, memory) to put mentalising knowledge to use, rather than testing individuals' knowledge of 'advanced' mentalising concepts.

The amount of research dedicated to exploring mentalising development from adolescence into adulthood is small in comparison to research on earlier years, but there is still evidence suggesting that individual differences in mentalising abilities can be observed in adulthood. Research on adult mental health often explores the influence of individuals' egocentric biases on decision-making, or individuals' tendency to rely too heavily on their own perspective (either visual or mental representations; Dumontheil et al., 2010). For example, research using referential communication tasks has shown that people make more errors in moving objects around a grid when they are instructed by a teacher who has a different visual perspective than their own (see Figure 1.5) (Zhao et al., 2018).

Figure 1. 5

Referential Communication Task (the Director's Task)



Note: In the director's task, the 'director' instructs the participant to move objects around a grid. However, the participant must bear in mind that the director has a different visual perspective than their own. Therefore, the director will only instruct the participant to move objects that they can see. In the experimental trial above, the participant must ignore their own egocentric tendency to move the largest cup they can see (distractor) and instead move the largest cup the director can see (the white cup in the bottom left corner).

Findings from studies using the director's task indicate that although individuals' ability to inhibit their own egocentric biases improves from young childhood to adulthood, adults still make errors on this task. Hence, our own perspectives distort and bias our decision-making and prevent us from representing other people's viewpoints clearly (Apperly, 2012, Dumontheil et al., 2010). Again, language complexity appears to support performance on this task, as altering the complexity of the instructions affects performance (Wang et al., 2016). Wang et al. (2016) found that 8- and 10-year-old children made more egocentric errors on the director's task when the complexity of the language used in the director's instructions (e.g., "nudge the small ball" versus "nudge the small ball one slot up") was increased. This suggested that the children's ability to

use the director's perspective when moving the objects around the grid was constrained by the complexity of language used and suggests that mentalising ability is linked to individuals' ability to comprehend spoken language. Apperly (2021) proposes that there are three key reasons that mentalising improves from middle childhood into adulthood: (1) more developed language and executive control mean that we can start to understand more complex social information; (2) increased practice means that *some* mentalising skills may become automated (e.g., judging others' visual perspectives), (3) increased social experience (socialising with friends, reading novels) improves our ability to use our mentalising abilities effectively (Jackson, 2020; Mar et al., 2006; Samson et al., 2010).

Overall, the literature indicates that mentalising continues to develop from young childhood to adulthood. Older children begin to use their mentalising knowledge in a more accomplished manner, and this improvement continues into adulthood. Furthermore, mentalising appears to become more fragmented after the age of five years old and the ability to understand mental states can be operationalised with different subcomponents (e.g., visuospatial perspective taking, false belief understanding, pragmatic language understanding) (See Section 3.2.1 for detailed discussion). Therefore, when examining the relationship between mentalising and other skills, it is important to be clear about which subcomponents of mentalising are related to the phenomenon of interest.

1.5 A framework for the direct involvement of mentalising in reading comprehension

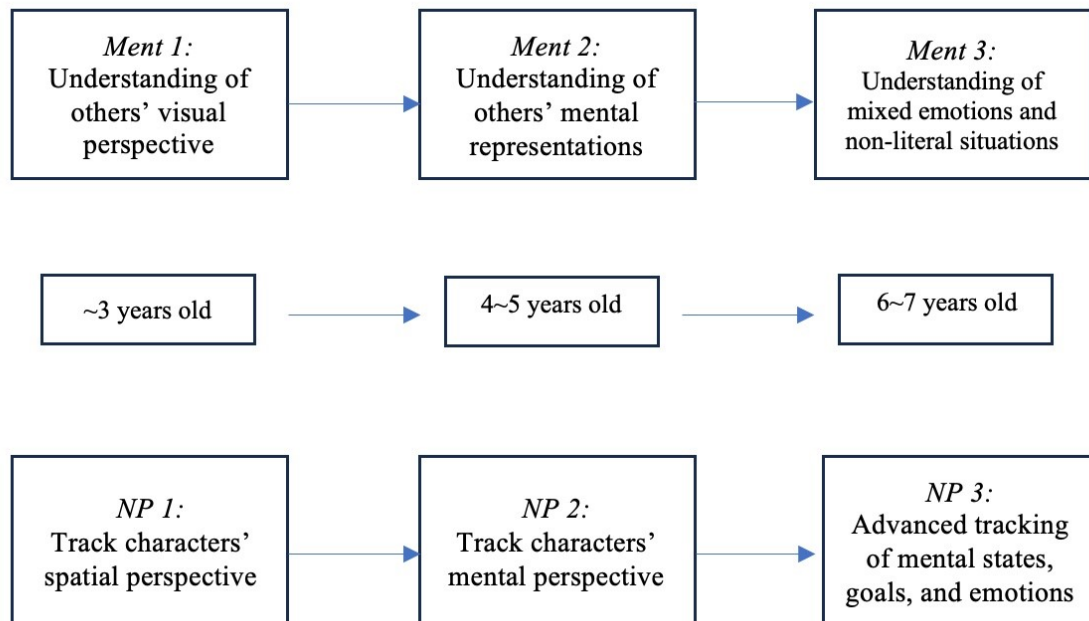
Dore et al. (2018) argue that mentalising may be the missing factor in explaining individual differences in reading comprehension. As narratives are mostly character-based, it is argued that without the ability to understand the mental states of characters, it is hard to understand the motivations behind characters' actions. Moreover, the motivations and desires of characters in fictional texts often propel the action in the narrative, so understanding the characters' perspectives may be key to understanding the story as a whole (Bruner, 1986). It follows that mentalising may support the comprehension of fiction texts specifically rather than expository texts. Dore et al. (2018) further argue that a relationship between mentalising and reading comprehension may be preceded by a relationship between mentalising and narrative processing in younger children. Indeed, children are first introduced to narrative stories orally, and mentalising may initially support their developing understanding of characters' thoughts, feelings, and drives when listening to narrative stories. As children learn to read independently, mentalising may aid the comprehension of narratives in fiction texts. Dore et al. (2018) propose that the qualitative stages of mentalising development from infancy to seven years old are closely aligned with the development of narrative processing during this period. Currently, although there is some experimental evidence of a relationship between mentalising and reading comprehension, there is little research that explains the mechanisms behind this relation (Atkinson et al., 2017; Kim 2016, 2017; Ebert, 2020). Therefore, Dore's framework indicates how mentalising may support reading comprehension through narrative processing and suggests testable predictions that may be investigated in future research.

1.5.1 The relationship between narrative processing and mentalising

Dore's framework asserts that the relationship between mentalising and reading comprehension can be traced to a pre-literate age when preschool children listen to children's stories. Children's stories, like adult literature, usually employ a narrative format in which the character's motivations and drives propel the narrative. At this young age, mentalising may support the child's understanding of the thoughts and feelings of characters and their actions, which helps them process the narrative. Dore et al. (2018) suggests that the development of narrative processing abilities is supported by a range of different mentalising abilities, such as visuospatial mentalising and mental state reasoning. The different aspects of mentalising that are proposed to support narrative processing in Dore's review may also support reading comprehension later in development.

Figure 1. 6

Alignment Between Mentalising Development & Children's Narrative Processing Abilities (Adapted from Dore et al., 2018)



Note. Ment = Mentalising, NP = Narrative processing.

Visuospatial mentalising may support narrative processing from the age of three years old since children have been shown to assume characters' visuospatial perspectives when listening to narrative stories. Rall and Harris (2000) read children's fairy tales (e.g., 'Little Red Riding Hood') in which verbs of motion (e.g., come, go) were used either consistently or inconsistently with the character's perspective. The children were then asked to accurately repeat the story back to the researchers. For example, the child would hear, 'Little Red Riding Hood was sitting in her bedroom when her mother went

in’; in this sentence, the verb of motion (went) is inconsistent with the Little Red Riding Hood’s perspective, whereas the verb ‘came in’ would be consistent with her perspective. The results indicated that when the verb of motion was consistent with the protagonist’s perspective, the children would accurately recall the sentence, including the verb of motion, but when the verb was inconsistent with the protagonist’s perspective, the children would often change the verb of motion to make it consistent (e.g., changing ‘went in’ to ‘came in’). When this study was replicated by Ziegler et al. (2005), they found that children were less likely to correct the verb if a non-human character (e.g., a toy car) was employed as the protagonist. The ability to track visuospatial perspectives in children under three years old is in line with research showing that infants are able to track the visual perspectives of agents from infancy (Kovács et al., 2010). Together, these studies suggest that rather than understanding the narrative from an external, third-person perspective, children adopt the internal perspective of the protagonist and may be more likely to take the internal perspective of an agentic character.

Dore et al. (2018) propose that when listening to narratives from the ages of four to five years old, children start to assume the characters’ mental perspectives and understand the inner motivations guiding their behaviour. O’Neil and Shultis (2007) examined development of the ability to take characters’ perspectives during a narrative in children from three to five years old. In this study, children were presented with a set of farmyard toys; the researcher read a narrative story about the toy figures in which the protagonist was thinking about a cow in a different location (a barn). When five-year-old children were asked to point to ‘the cow’, they pointed to the cow that the protagonist was thinking about, whereas when three-year-old children were asked to point to ‘the cow’, they pointed to the one in close proximity. These findings signal that the older

children were able to track the mental perspective of the protagonist, whereas the younger children were using the visuospatial perspective of the protagonist to infer which cow the researcher was referring to. The ability to take the protagonist's mental perspective therefore appears to develop after three years. In addition, further studies have indicated that four-year-old readers may be able to understand that characters can have subjective attitudes toward objects and events. Fecica and O'Neill (2010) found that four-year-old children processed 'horrible' events slower than pleasurable events when reading, suggesting that the listener was adopting the characters' subjective feelings. Together, these results highlight that young children adopt the mental states and subjective feelings of characters when listening to stories. Importantly, this development in narrative processing occurs at a crucial stage of mentalising development, namely when children are able to see that other people may hold inaccurate representations (e.g., false beliefs) of reality (Astington & Jenkins, 1999; Rakoczy, 2022).

Dore et al. (2018) further propose that when children reach six to seven years old, they start to interpret characters' goals and motivations which drive the narrative forward. Filiatrault-Veilleux et al. (2016) investigated the ability of children from three to six years old to infer the mental states of characters and predict subsequent behaviour. In this study, children looked at illustrations of a story while listening to the narrative. After each illustration, an inferential question was asked. At four to five years old, children were able to infer the mental states of characters and predict the next steps of the story but could not interpret characters' goals. However, from five to six years old, children were better at making inferences about the goals of the protagonist and using this information to predict the resolution of the story. Moreover, Nyout (2015) designed a task in which six- to seven-year-old children listened to a short story where the character's goal (e.g., to read a book) was stated at the beginning. Processing times were

longer during reading when a later statement was incongruent with the original goal, implying that children spontaneously monitor characters' goals throughout narrative processing. Together, these studies suggest that by ages six to seven, children are able to identify the motivations of characters and use this information to understand the narrative action.

The ability to understand the motivations of characters in narratives occurs at the same age that children begin to understand advanced mentalising concepts (e.g., white lies, double bluffs; O'Hare et al., 2009). Dore et al. highlight that in comparison to research in children from three to five years old, there is a lack of research in either narrative processing or mentalising to trace a clear alignment between the abilities at this age. Despite the lack of research at this age, previous research would suggest that the reason a lack of coherence is found between mentalising and narrative processing at this age is because there is a lack of coherence between mentalising tasks after five years old (Osterhaus & Bosacki, 2022). Therefore, it is harder to correlate narrative processing skills with mentalising as a single concept, and it may be necessary to correlate mentalising with different subcomponents of mentalising. For example, recent research has found a relationship between mentalising (specifically, mental state reasoning, measured by the Strange Stories) and narrative comprehension in both 11 to 15-year-old autistic and neurotypical students (Harvey et al., 2024). Therefore, there may be a special relationship between mental state reasoning and narrative processing, rather than general mentalising abilities and narrative processing skills.

Overall, Dore et al.'s (2018) review describes a conceptual alignment between the development of children's narrative processing skills and mentalising abilities from three to seven years of age. Mentalising may support the development of narrative processing during this period by aiding in the representation and monitoring of

characters' perspectives when listening to narrative stories. A key point illustrated by Dore's framework is that there are several subcomponents of mentalising (taking a character's spatial perspective, then their mental perspective, then interpreting their goals), and these subcomponents of mentalising may be related to reading comprehension in different ways. In summary, Dore's evidence of the temporal alignment between mentalising and narrative processing does not provide evidence of a causal link between the concepts, but it does indicate that further research is warranted to explore the relationship.

1.5.2 Approaches to studying reading comprehension

When examining Dore et al.'s (2018) temporal alignment of the development of mentalising and narrative processing abilities in childhood, the link between the two concepts is weakest during the last stage (six to seven years old) due to the authors' conceptualisation of mentalising. In their review, they approach mentalising as a set of milestones of conceptual development that span from early to middle childhood (see Section 1.2). Consequently, during the two stages where mentalising milestones are clearly described (three to five years old), it is easier to link the age at which children understand these mentalising concepts with narrative processing abilities (O'Hare et al., 2009; Wellman & Liu, 2004). However, after the age of five, there are no clearly demarcated mentalising milestones to link with narrative processing developments (Apperly, 2012). Dore et al. (2018) acknowledge that the last stage of alignment between mentalising and narrative processing abilities is based on a smaller body of empirical

evidence. Therefore, the conceptual view of mentalising is challenging for researchers studying the influence of mentalising in middle childhood or beyond.

In contrast, the cognitive approach to mentalising describes how mentalising develops continuously after the age of five and is more suited to explaining how mentalising supports reading comprehension in adulthood. This approach proposes that, after critical mentalising milestones are reached in early childhood, the ability to use mental state reasoning continues to improve, but this improvement is related to improvements in other cognitive domains (e.g., language, working memory). Thus, the cognitive approach to mentalising may be more useful in explaining how a child develops from understanding single mental state concepts in simple stories (e.g., Sally Anne Task) to understanding a range of complex, interacting mental states across characters in a fictional narrative in adulthood. Therefore, from a cognitive perspective it is important to control for abilities that may facilitate mentalising (e.g., language abilities, working memory) in order to measure the effect of mentalising in adults.

1.5.3 Mentalising may aid the creation of situational models (Kintsch, 1988)

Dore et al.'s (2018) framework explains how mentalising supports the construction of situational models during reading and why mentalising has a direct effect on reading comprehension. Like the DIET and DIER models, Dore et al. use Kintsch's (1988) Construction Integration Model as a theoretical framework to explain how mentalising may be employed for reading comprehension. Here, the highest level of comprehension is the situational model – i.e., a mental representation of key events, characters, and plotlines constructed by the reader. The situational model is constructed

from explicit information in the narrative, as well as inferences made by integrating explicit information with the reader's own background knowledge. Mentalising may support the construction of rich situational models by helping readers make more accurate inferences about characters' mental states.

Dore et al. (2018) propose that mentalising will differentially predict texts depending on whether they include character-based properties. Accordingly, their framework presents a number of testable hypotheses. Firstly, mentalising will be drawn upon more when comprehending narrative texts that involve characters and less when comprehending expository texts. Although this is a logical proposition based on their framework, the empirical research in young children that has found a relationship between mentalising and reading comprehension does not necessarily support this assertion. Kim (2017) found that there were moderate correlations between mentalising and narrative comprehension ($r = .51$) and mentalising expository text comprehension ($r = .47$), suggesting that mentalising has a role in comprehending both types of text. In many studies, it is harder to separate narrative from expository texts in past research; for example, Kim (2017), Ebert (2020), and Atkinson et al. (2017), the reading comprehension measures used a mixture of fictional and non-fictional narratives. It is, therefore, possible that there were stronger associations between mentalising and fictional texts in these studies but that these effects were masked as they used combined reading measures with both types of texts (Snowling et al. 2011).

On the other hand, there are reasons to expect that mentalising may not be differentially related to fictional vs expository texts in adults. Although mentalising may be recruited more in literary texts than expository texts, it may be hard to elicit this effect in adults as better adult readers tend to be good at comprehending both types of texts. This may be because proficient adult readers read more pleasure at a young age, which

may be encouraged in the home literacy environment. Reading for pleasure has been associated with improvements in comprehension of both fictional and expository texts throughout the lifespan, (Boerma et al., 2017; Mol & Bus, 2011). In Torppa et al.'s (2022) longitudinal study, the effects of the home literacy environment on literacy levels were measured from birth to 23 years old. The authors found that shared reading between children and their parents at two years old predicted vocabulary development and reading motivation and that reading motivation was a strong predictor of reading for pleasure at 15 years old. There is evidence that mentalising may facilitate the relationship between the home literacy environment and reading comprehension. Boerma et al.'s (2017) study examining middle childhood (8-11 years old) found that mentalising mediated the relationship between the home literary environment (e.g., literacy resources, reading at home) and reading comprehension. Boerma et al. (2017) suggests that as children grow older, they may be given more complex texts that require more inferencing about the mental states of characters, drawing on mentalising skills (Norbury & Nation, 2011). In summary, the home literacy environment appears to motivate children to read for pleasure, and increased reading for pleasure improves children's reading comprehension, particularly in children with better mentalising abilities.

Furthermore, there may be a bidirectional relationship between mentalising and reading (Jackson, 2022). Mentalising may help readers understand the mental states and emotions of characters, which helps readers generate richer situational models of the text; in turn, fictional narratives may give readers a simulated experience of social interactions between characters, which improves mentalising skills (Mar & Oatley, 2008). However, so far, there is only sparse correlational evidence supporting this bidirectional relationship (Jackson, 2020; Mar et al., 2006). In summary, mentalising may

support reading fictional texts, and this relationship may be bi-directional. However, it is challenging to find evidence of this relationship as individuals with better comprehension (possibly aided by mentalising) tend to read more, which improves their reading comprehension of different text types (fiction and expository).

A second prediction in Dore et al.'s framework is that mentalising should support inferences about a character's cognitive and emotional states but have weaker associations with literal statements and causal inferences during reading. Therefore, Dore proposes that performance on mentalising tasks will predict performance on inferencing tasks that draw upon mental states but not tasks that draw on background knowledge unrelated to mental states. In Graesser et al.'s (1994) taxonomy of inferences, this would mean that mentalising supports inferences about characters' superordinate goals (i.e., goals that motivate a character's intentional action) and inferences about characters' emotional reactions. For example, when inferring a character's emotional response (e.g., sadness, anger) to an event or action (e.g., losing all their money while gambling). However, according to this framework, mentalising would not support causal inferences (e.g., inferring how much money a character won from a bet).

To explore the relationship between mentalising and reading comprehension in more detail, as Dore recommends, it would be useful to examine the fine-grained relationships between mentalising and specific reading comprehension sub-skills, such as mental state inferencing. Measuring mental state inferences may involve recording readers' ability to make inferences about characters' mental states and the reasons for their actions. For example, if mentalising supports inference generation about characters' mental states, it may have particular importance when making pragmatic inferences about the meaning of characters' in-text dialogue. Pragmatic inferences are needed when the intended meaning of a character's spoken dialogue goes beyond the

literal meaning of words (Spotorno et al., 2016). For example, during a narrative dialogue between two characters, Peter says, ‘Would you like to visit my sister?’ John replies, ‘She was incredibly rude to me last time we met.’ Here, the reader must go beyond the literal meaning of the reply to infer that John does *not* want to visit Peter’s sister. In this instance, the literal meaning of the sentence (Peter’s sister was rude) underdetermines the speaker’s intended meaning (he does not want to visit her). However, there might be information in the preceding narrative context that further alters the meaning of the utterance. For example, John might be in love with Peter’s sister and want to see her, despite her being rude the last time they met (Frank, 2018). By monitoring the characters’ perspectives within a rich situational model, the reader may be better able to make accurate inferences about the meaning of their dialogue even when the character’s intended meaning is not explicitly stated. Hence, it is possible that mentalising supports pragmatic inference in a similar way to mental state inferences.

In summary, Dore et al.’s (2018) framework postulates that better mentalisers will be better at representing and monitoring explicit mental state information during reading comprehension. This enables them to make more accurate inferences about characters’ mental states when this information is not explicitly mentioned. This, in turn, aids in creating rich and coherent situational models during reading. Mentalising should support readers’ inferences about characters’ mental states, but not literal comprehension or causal inference generation. Currently, there is little empirical evidence that clearly shows a link between mentalising and narrative texts specifically (as opposed to expository texts), as most researchers have used a mix of literary and expository texts to measure reading comprehension (Atkinson et al., 2017; Ebert, 2020; Kim, 2017). Therefore, testing whether there is a relationship between mentalising and fictional texts specifically would be the next logical step in testing this framework.

1.6 Chapter Summary

In summary, the research examined in this chapter demonstrates that reading comprehension is a complex process that draws on multiple sub-skills and the way in which it continues to develop beyond childhood has not yet been fully characterised. The SVR suggests that reading comprehension is a product of two components: word decoding and oral language (Gough & Tunmer, 1986; Hoover & Gough, 1990). However, criticisms of this model are (1) that it does not adequately describe the structure of different skills within each of these components, and (2) that it does not describe how these components interact during the process of reading.

To address these issues, Kim (2017) expands on the SVR by unpacking the oral language comprehension component, describing the structure of lower-level skills (e.g., working memory, vocabulary) and higher-level cognitive abilities (e.g., inferencing, mentalising). Additionally, to trace the process of reading comprehension, Kim describes the way in which comprehension processes operate at the lexical, sentence, and discourse levels. Kim identifies mentalising as a higher-order sub-skill that enables readers to reason about the mental states of characters in narratives and to create rich situational models of the text. In Kim's (2016, 2017) empirical studies, mentalising is indirectly related to reading comprehension via oral language comprehension in children. As oral language comprehension accounts for more variance in reading comprehension as we age, the influence of mentalising on reading comprehension will increase in importance with age.

Further, Dore et al. (2018) propose that mentalising may have a special function in supporting comprehension of fictional texts that stems from an analogous

developmental trajectory of mentalising and narrative processing at a preliterate age. Dore shows that from the ages of three to seven years old children develop an understanding of different subcomponents of mentalising (visuospatial perspective taking, mental state reasoning, and goal tracking). By the age of seven, when they have acquired skill in these different subcomponents of mentalising, they can establish a more sophisticated understanding of characters' character mental states in narrative fiction. Therefore, it is important to examine the relationships between different subcomponents of mentalising and reading comprehension, as subcomponents may support adult reading comprehension differentially. The key mechanism proposed is that mentalising supports both the representation and monitoring of explicit information about characters' mental states in the text and accurate mental state inference generation, thus supporting a rich situational model in narrative text reading.

Although there is evidence that mentalising may have an independent role in reading comprehension, the evidence base supporting this claim is small and mainly based on research in young children (Atkinson et al., 2017; Boerma et al., 2017; Kim, 2016, 2017). It is possible that a relationship between mentalising and reading comprehension only exists in young children, related to the development of conceptual mental state milestones in early childhood. Alternatively, it may be that individual differences in readers' ability to *use* their knowledge of mental state concepts to understand characters' mental states in fictional texts is related to their reading comprehension. If this is the case, mentalising may continue to support reading comprehension into adulthood (via language comprehension). Furthermore, as research shows that language comprehension has an increasingly important role in reading comprehension across development after the age of 10, the role of mentalising in reading

comprehension may also increase with development (Garcia & Cain, 2014; Massonnié et al., 2018).

1.7 Aims and research questions

This thesis aims to investigate the nature of the relationship between mentalising and reading comprehension in adults and to explore potential mechanisms underpinning this relationship.

The research reported in this thesis aims to address four overarching research questions.

1. What is the relationship between mentalising and reading comprehension in adults?
2. What subcomponents of mentalising do current tasks used with adults measure?
3. How are these subcomponents differentially related to reading comprehension in adults?
4. What mechanisms underpin the relationship between mentalising and reading comprehension in adults?

2. Investigating the Relationship Between Mentalising and Reading Comprehension in Adults: A Correlational Study (Study 1)

2.1 Chapter overview

Chapter 1 provided a theoretical framework describing how mentalising may support reading comprehension. However, most existing research is focused on young children, with little research examining the role of mentalising in adults (e.g., Kim, 2016, 2017). It is possible that a relationship between mentalising and reading comprehension may only be found in young children as the conceptual milestones in the early theory of mind understanding are reached (e.g., Atkinson et al., 2017). On the other hand, reading comprehension may be related to individuals' ability to *use* their knowledge of mental state concepts flexibly to understand social phenomena; if this is the case, then it is likely that mentalising will support comprehension of complex

narratives across the lifespan, as the ability to deploy conceptual knowledge develops from early childhood to later adulthood (Apperly, 2012; Carlson & Moses, 2001; Cavallini et al., 2013; Li et al., 2013).

2.2 Introduction

Concurrent and longitudinal associations between theory of mind and reading comprehension in young children was demonstrated by Atkinson et al. (2017). Specifically, in this study false belief understanding at age three predicted reading comprehension two years later both directly and indirectly (via oral language comprehension). Although the two false belief tasks used in this study placed demands on children's verbal skills presenting a potential confound, the study design controlled for skills related to reading, such as verbal abilities (e.g., word decoding, vocabulary) and executive functions (e.g., working memory; Hogrefe et al. 1986). These findings suggest that the *knowledge* of false beliefs is associated with reading comprehension in young children.

There is a small amount of research that suggests that the ability to *use* mentalising concepts in a flexible manner may continue to support reading comprehension later in development (Boerma et al., 2017; Ricketts et al., 2013). Boerma et al. (2017) determined that mentalising mediated the relationship between the home literacy environment and reading comprehension in middle childhood (8-11 years old). The home literacy environment refers to a range of parental attitudes towards reading, the availability of literacy resources, and reading activities at home (Park, 2008). Boerma et al.'s (2017) results suggest that children's engagement with books at home may support the development of advanced mentalising skills, which in turn support

reading comprehension, although it is important to note that the concurrent design limits causal inference. The measure of mentalising in this study was the Strange Stories (White et al., 2008), a text-based mental state reasoning task that draws on people's ability to interpret the mental states of people in social situations drawing on pragmatic language skills (e.g. interpretation of white lies, double bluff). Thus, there is a potential overlap in task demands between the Strange Stories and standard measures of reading comprehension, which raises questions as to the specific contribution of mentalising. Further, while children's *expressive* language skills were included in the model, the study design did not control for important sub-skills within the SVR (i.e., word decoding and oral language *comprehension*). Whether mentalising is a unique predictor of reading comprehension in middle childhood or is better viewed as a sub-component of oral language comprehension (Kim, 2017), therefore, remains unclear.

A concurrent predictive relationship between mentalising and reading comprehension has also been reported in autistic adolescents (average age 15 years old; Ricketts et al., 2013). After controlling for the components of the SVR (word decoding and oral language comprehension), mentalising predicted unique variance in reading comprehension. To compare which subcomponents of mentalising may be related to reading comprehension, they used two mentalising tasks: the Strange Stories and a visual mentalising task, the Frith-Happé Animations (Castelli et al., 2002). In this task, participants are asked to ascribe intentions to the movements of animated shapes. Both mentalising tasks predicted significant variance in reading comprehension. Autistic children and adults have been shown to score more poorly on mentalising tasks than non-autistic peers on average in a number of studies (e.g., Brewer et al., 2017; Schwartz et al., 2022). Therefore, on the one hand, it may be that mentalising ability has a particularly strong impact on reading comprehension in autism. On the other hand,

substantial variability in mentalising ability has also been observed in typically developing children and adults (Cavallini et al., 2013; Hughes & Devine, 2015; Wang & Su, 2006). Therefore, it may be individuals with or without autism, who have lower mentalising abilities consequently have problems in reading comprehension.

Correlational evidence supports the persistence of a relationship between mentalising and reading comprehension into adulthood. Jackson (2022) reported correlations between mentalising, measured using the Reading the Mind in the Eyes Test (Baren Cohen et al., 2002), and a standardised measure of reading comprehension, the York Adult Assessment Battery-Revised (YAA-R; Warmington et al. 2013). In this study, the Reading the Mind Task was used rather than a text-based measure of mentalising to reduce the load on reading comprehension. In Reading the Mind in the Eyes Test, participants are shown isolated pictures of eyes and asked to choose a matching cognitive or emotional state (e.g. thinking about something; bored) from a choice of four. This task therefore has fewer overlapping demands with measures of reading comprehension than text-based measures of mentalising. Jackson's (2022) results suggest that the relationship between mentalising and reading comprehension is not specific to mentalising as measured by the Strange Stories.

Finally, a recent meta-analysis by Tompkins et al. (2024) examined the relationship between mentalising and text comprehension across the lifespan. In this meta-analysis, 47 independent samples with a total of 5,123 participants aged 3 to 70 (*mean age* = 10.53 years) were analysed. The study found a significant association ($r = .33$) between mentalising abilities and text comprehension. This relationship remained significant even after controlling for verbal and executive function skills, with effect sizes ranging from, $r = .22-.32$. The strength of the association did not vary based on factors such as study design (cross-sectional or longitudinal), participant age, or

specific characteristics of the or comprehension tasks. Notably, the study found that relationship between mentalising and text comprehension was found across different mentalising tasks drawing upon different mentalising abilities (e.g., false belief understanding, making attributions from non-verbal cues, mental state reasoning). These findings support a link between mentalising and text comprehension throughout different stages of life.

Therefore, there is correlational evidence to demonstrate a robust association between mentalising and reading comprehension beyond the conceptual theory of mind milestones of early childhood, through late childhood and into adulthood. At present, this evidence is largely correlational and cross-sectional, which constrains causal inference. The studies reviewed here also indicate that different indices of mentalising are related to reading comprehension, including mental state reasoning from short texts (the Strange Stories), attributing mental states based on non-verbal cues (Reading the Mind in the Eyes), and attributing intentions to moving shapes (Frith-Happé Animations).

Mentalising may continue to support the comprehension of narrative texts in older age. However, the results of the few studies that have examined mentalising ability across the lifespan are mixed on whether mentalising remains robust or declines in older age. Rokoczy et al. (2017) found that mentalising declined in older adults (61 to 78 years old) in comparison to younger adults (18 to 30 years old). However, this study did not control for inhibition and working memory, which have been found to mediate the relationship between age and mentalising (Li et al., 2013; Long et al., 2018). Long et al. (2018) found that older adults (< 65 years old) with higher inhibition scores still scored highly on mentalising tasks. On the other hand, it may depend on the type of mentalising task being measured. Bottiroli et al. (2016) found that working memory mediated the

relationship between age and cognitive mentalising (making inferences about thoughts and beliefs) but that affective mentalising (making inferences about emotions) remained robust in older age and was not reliant on other skills. In summary, although the results are mixed, evidence suggests that mentalising remains robust in adults with uncompromised cognitive abilities in older age.

If unimpaired, mentalising may continue to support reading comprehension into older adulthood, particularly in narrative texts. In narrative texts, characters' motivations are intrinsically entwined with the events in the text. Mentalising may support understanding of characters' mental states and improve comprehension of the text as a whole. On the other hand, mentalising may not support comprehension when reading expository texts because these texts are largely factual and less reliant on characters' mental states to drive a plot. De Beni et al. (2007) reported a decline in expository text comprehension in adults over 65 years, whereas narrative text comprehension is relatively preserved. De Beni et al. (2007) compared older adults (64-74 years old) and younger adults (18-30 years old) reading comprehension ability on narrative and expository texts. Here, the narrative text required participants to answer comprehension questions about a conflict between an adolescent and their parents, whereas the expository text required comprehension of Japanese adults' television watching habits. Older participants' narrative text comprehension was equivalent to young participants' narrative text comprehension, but the older adults' expository comprehension was impaired. This decline in expository text comprehension was related to cognitive decline in working memory and metacognitive flexibility. In contrast, older adults were able to use other mental processes, such as the integration of background knowledge, to understand the narrative text. Lin et al. (2000) have also shown that using metacognitive strategies, such as identifying key pieces of information essential for understanding, has

been shown to predict reading comprehension of narrative texts. Since the characters' motivations drive the plot of narrative fiction (Bruner, 1987), older readers may also recruit preserved mentalising ability to infer characters' mental states, supporting narrative text. However, thus far, researchers have not explored the relationship between mentalising and reading comprehension across the lifespan.

One obstacle to researching mentalising and reading comprehension in adults is that there are few standardised tasks designed to measure either of these two constructs beyond childhood. In addition, past research also suggests that different mentalising measures show minimal coherence with each other, meaning that performance on one mentalising task does not predict performance on another measuring the same ability (Warnell & Redcay, 2019). This is problematic as past research that has associated mentalising with aspects of reading and language has used a variety of different tasks that come under the umbrella term 'mentalising', but may in fact measure minimally related sub-components. For example, improved performance on the mentalising task, the Reading the Mind in the Eyes Test, has been associated with reading literary fiction in adults (Baron-Cohen et al., 2002; Kidd & Castano, 2013). In middle childhood, mentalising, as measured by the Strange Stories, has been shown to mediate the relationship between the home literary environment and reading comprehension in middle childhood (Boerma et al., 2013). Yet if the Reading the Mind in the Eyes Test and the Strange Stories are in fact not measuring the same construct, but rather specific subcomponents of mentalising (e.g., the ability to attribute emotions to pictures of eyes and interpreting pragmatic language, respectively), then the role of mentalising in reading comprehension becomes less clear. Therefore, it is important to specify which subcomponents of mentalising are being measured rather than using the term mentalising as a catch-all descriptor of diverse cognitive processes.

Over the last 30 years, researchers have produced a range of tasks to measure mentalising that differ across a number of dimensions, such as modularity (visual vs verbal stimuli), and whether they test implicit or explicit mentalising, cognitive and affective processes. As these mentalising tasks have been developed, the way in which mentalising is understood theoretically has changed (See Section 3.2). Many commonly used mentalising tasks (e.g., Strange Stories (Happe, 1994) and the Faux Pas Task (Stone & Baron-Cohen, 1998)) were designed when mentalising was generally conceptualised as a unidimensional construct working independently from other cognitive skills (e.g., working memory) and controlled by a specific neural network (Frith & Frith, 2006; Leslie & Thaiss, 1992; Saxe, 2006). Yet more recent researchers have questioned this conceptualisation. Gerrans and Stone (2008) suggest that rather than existing as a domain-specific module, mentalising may be better understood as a multidimensional construct that is made up of multiple low-level domain-specific social processes (e.g., visuospatial mentalising, interpreting pragmatic language) and supported by other domain-general abilities (e.g., executive functions). In support of this, recent neuroscientific research has questioned whether mentalising is controlled by a single mentalising network, instead, different subcomponents of mentalising may be controlled by distinct areas of the brain (Schaafsma et al., 2015). The lack of coherence between tasks after the age at which children understand false beliefs (four to six years old) further indicates that mentalising is not a unidimensional construct (Apperly, 2012). Moreover, developmental improvements in the ability to use mental state knowledge to understand others are closely related to the development of domain-general abilities.

To assess whether mentalising tasks show coherence with one another (which would suggest a unidimensional construct) or not (which would suggest a multidimensional construct), Warnell and Redcay (2019) examined patterns of

association between mentalising tasks across three different age groups. Mentalising batteries comprised tasks tapping a range of sub-skills (e.g., attributions from non-verbal cues, interpreting pragmatic language) and drew on different non-mentalising cognitive skills (e.g., visual processing, language comprehension). Previous studies that have reported strong associations have tended to measure a small number of mentalising tasks. For example, Devine and Hughes (2016) found coherence between a text-based measure, the Strange Stories, and a film-based mentalising task drawing on visual processing, the Silent Films task. The latter was designed to measure the same advanced mentalising concepts as the Strange Stories without drawing on reading comprehension. However, these tasks may have shown coherence as they measure the same subcomponent of mentalising (i.e., mental state reasoning). Warnell and Redcay (2019) aimed to explore the coherence of a wider range of tasks measuring various sub-components of mentalising (e.g., attributing mental states to pictures of eyes vs. understanding social transgressions). Additionally, to investigate whether the structure of mentalising varied across development, they examined performance in early childhood (four- to six-year-olds), middle childhood (seven to twelve years old), and adults (average age = 20 years old). Where possible, they used tasks that examined similar mentalising sub-skills abilities in each age group; for example, all groups completed the Reading the Mind in the Eyes Test. In early childhood, the children were given false belief tasks, Simplified Reading the Mind in the Eyes Test, an appearance reality emotion task (in which children had to listen to a story and describe the discrepancy between the character's actions and their true feelings), and the Faux Pas Task (in which the child must identify violations of social norms). In the middle childhood group, they were given the Reading the Mind in the Eyes Test, the Strange Stories, and the Faux Pas Task. In the adult group, participants took the Spontaneous Theory of Mind Protocol

(STOMP; Rice & Redcay, 2015), in which participants viewed two silent movie clips and then described what happened in the clip, the Belief-Desires task (Apperly, 2011), which is a timed false belief task, the Pragmatic Language Comprehension task (Koster-Hale, Dodell-Feder, & Saxe, unpublished), in which they joined sentences in appropriate order, the Reading the Mind in the Eyes Test, a higher order belief task (Rice & Redcay, 2015) in which participants listened to stories and answered questions about the characters beliefs. Findings demonstrated coherence between conceptually similar mentalising tasks (e.g., false belief tasks) in early childhood (Liu et al., 2008; Wellman et al., 2001). However, after this age, there was little coherence between the different tasks after controlling for variables such as age and verbal abilities. Warnell and Redcay (2019) conclude that the lack of coherence between the mentalising tasks reflects the multi-dimensional nature of mentalising as a construct. They also suggest that, if mentalising is a multidimensional construct supported by different cognitive abilities, then mentalising tasks should be deconstructed to understand which subcomponents (e.g., visuospatial perspective-taking, interpreting social situations) and which non-mentalising cognitive abilities (e.g., language abilities, executive functions) they draw upon (Apperly, 2012; Schaafsma et al., 2015).

However, there may be other practical and theoretical reasons why a lack of coherence between mentalising tasks was observed in this study. Osterhaus and Bosacki (2022) argue that some tasks are better measures of developmental differences than individual differences, so comparisons between the tasks may not be suitable. For instance, in Warnell and Redcay's (2019) analysis, the Faux Pas task did not show coherence with other mentalising tasks (Reading the Mind in the Eyes, False belief task, visual perspective taking task) within the preschool children group, but it was sensitive to developmental differences between the 6-year-olds in the preschool group and the

seven- to twelve-year-olds in the middle childhood group. Another issue when comparing different mentalising tasks is that their reliability is often not reported (Osterhaus & Bosacki, 2022), and when psychometric properties have been investigated, reliability has been found to be low (Hayward & Homer, 2017). Therefore, the lack of coherence between mentalising tasks reflects measurement error rather than conceptual differences.

Understanding the properties of tasks is particularly important when investigating the role of mentalising in reading comprehension, as many available tests of adult mentalising rely heavily on language comprehension (e.g. Strange Stories), meaning that there will be several overlapping task demands with measures of reading. For instance, the reader may need to use vocabulary and inferencing skills to monitor and update their knowledge of characters' mental states as they read (Apperly, 2018; Dore et al., 2018). Therefore, these variables must be controlled to disentangle any unique contribution of mentalising to reading comprehension above and beyond shared sub-skills such as vocabulary and inference generation. Additionally, it is important to use multiple measures of mentalising as different subcomponents have been associated with reading abilities. For example, mental state reasoning, visuospatial perspective-taking and attributions from non-verbal cues have all been associated with reading abilities in previous studies (Boerma et al., 2017; Kidd, & Castano, 2013; Zwaan & Radvansky, 1998). Currently, there are few validated measures of mentalising designed specifically for adult. Therefore, it may be necessary to use tasks, e.g., the Strange Stories, that were not designed for neurotypical adults but have been shown to measure variability in this population, in order to find task to measure different subcomponents of mentalising (Wang & Su, 2006).

Similarly, although there are a small number of reading comprehension tests standardised in adult populations in the United Kingdom that use expository texts (e.g., the Adult Reading Test, 2nd Edition [ART-2]; Everatt et al., 2017), there are currently none that use narrative texts. This study, therefore, combines the ART-2 with the reading comprehension subsection of the Short Story Task (Dodell-Feder et al., 2013) to create indices of reading comprehension in adults that cover both expository and narrative texts.

In summary, there is limited research evaluating the relationship between mentalising and reading comprehension in adults. An obstacle to conducting research in this field is that there are few measures of either construct designed for adults. The first aim of the current study is, therefore, to examine the correlations between four different mentalising measures in an adult sample to assess coherence between the tasks. Based on previous research, it is expected that there will be weak or minimal correlations between the different mentalising measures (Warnell & Redcay, 2019). A second aim is to examine whether mentalising predicts reading comprehension in adults and, if so, which subcomponents of mentalising are significant predictors. In line with childhood research (Kim 2017, 2021; Kim et al., 2021), it is expected that mentalising will predict reading comprehension after controlling for vocabulary and inferencing. Also, a secondary prediction is that that mentalising will be a stronger predictor of comprehension of narrative texts than expository texts (Dore et al., 2018).

2.3 Method

Participants

A power analysis was conducted using G-Power for a multiple regression yielding a statistical power of at least .80 with an alpha of .05 and a medium effect size ($d = 0.50$). Using these parameters, the minimum recommended sample size was 85. Participants were recruited online via two platforms: Prolific and SONA, the York St. John University online participant recruitment service. The main difference between these recruitment platforms is the population that each recruits from and the way in which participants are rewarded. Prolific recruits participants from the general public, who are paid for their participation. Participants recruited via SONA, on the other hand, are exclusively from York St. John University Psychology taught undergraduate and postgraduate programmes and receive course credit. There were four pre-screening criteria: (1) age (participants had to be over 18); (2) nationality (participants had to come from the United Kingdom or Ireland); (3) first language (participants had to speak English as a first language); (4) no literacy difficulties, as determined by asking participants the following question: ‘Have you been diagnosed with dyslexia, dyspraxia, or ADHD (attention deficit hyperactivity disorder), or you aware of having any related reading difficulties?’ Fifty-two participants were recruited through Prolific and were paid £7.50 for participating. Two participants’ submissions were rejected as their data suggested they had repeatedly failed to follow task instructions, leaving a total of 50 participants. Forty-seven participants were recruited via SONA, the York St. John University online recruitment service. Participants were all undergraduates who gained course credits for their participation. Two participants were removed for failing more

than three instructional manipulation checks (IMCs), and one participant was rejected for consistently not following task instructions. This left 44 participants recruited via SONA. Therefore, 94 full participant datasets were used for analysis. Participant demographic characteristics are summarised in Table 2.1.

Table 2. 1

Demographic Details

Platform	Number of participants	Age Range (mean, <i>SD</i>)	Gender
Prolific	50	18–68 (40.18, 12.76)	22 females, 28 males
SONA	44	18–31 (19.84, 2.95)	44 females
Whole Sample	94	18–68 (30.70, 13.92)	66 females, 28 males

Research design

This study employed a correlational, cross-sectional design using multiple indicators of two psychological concepts (mentalising and reading comprehension) while controlling for aspects of language ability (vocabulary and inferencing skills). Mentalising was the key predictor variable of interest. It was operationalised using the Director’s Task (visuospatial perspective-taking), Reading the Mind in the Eyes Test (attributions from non-verbal cues), Strange Stories (mental state reasoning), and the Short Story Task (interpreting social situations within an extended narrative text). The outcome variable was reading comprehension, operationalised using the ART-2 (a standardised measure of reading comprehension using expository texts), and the Short

Story Task reading comprehension subsection (an unstandardised measure of reading comprehension using narrative texts).

Two aspects of language ability were included as control variables: vocabulary knowledge (operationalised using the LexTALE task) and inference generation (operationalised using the Implicit Management Task).

Materials

The materials for this experiment were adapted for online use using the software programme Gorilla.

Measures of mentalising

Visuospatial perspective-taking (The Director's Task)

The Director's Task (Dumontheil, et al., 2010) is a test of visuospatial perspective taking. The participant listens to instructions given by the director, who asks the participant to move one of the eight objects from one box to another (see Figure 2.1). There were 48 trials in total (8 experimental trials, 8 control trials, and 32 filler trials). The Director's Task is calculated by subtracting the percentage of experimental errors from the percentage of errors on the control task. Internal reliability in the current study was high (Cronbach's $\alpha = .89$).

Figure 2. 1

The Display Screen in the Director's Task

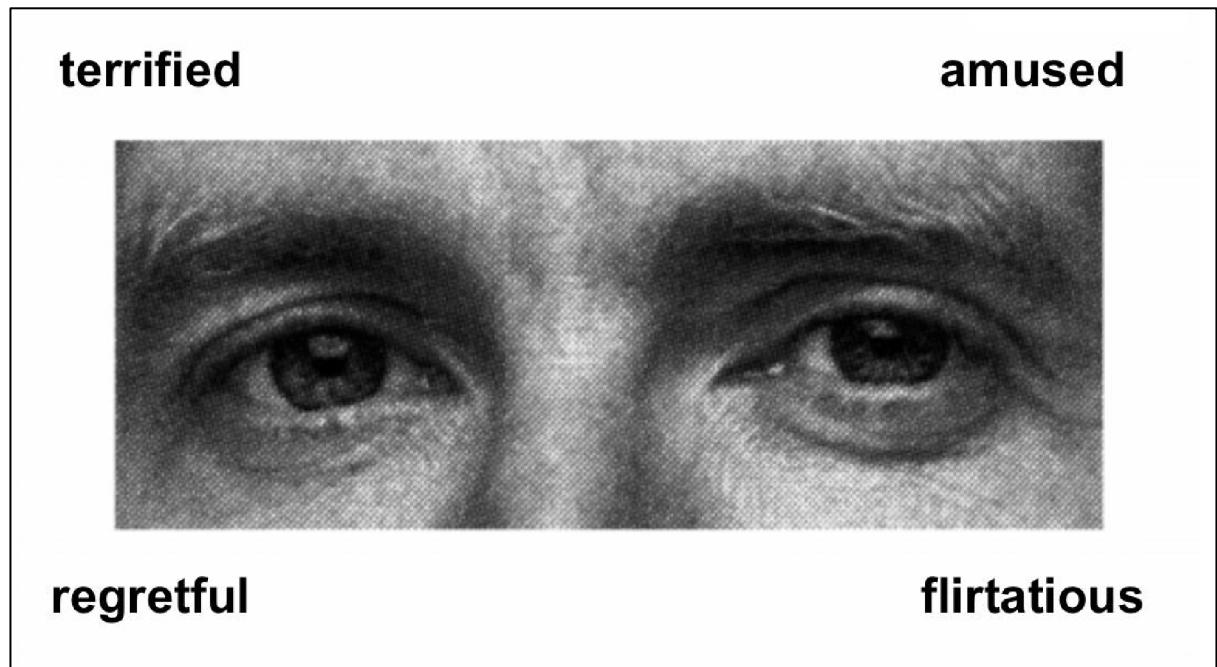


Attributions from non-verbal cues (Reading the Mind in the Eyes Test)

The Reading the Mind in the Eyes Test (Baron-Cohen et al., 2001) measures individuals' ability to make attributions from non-verbal cues (see Figure 2.1). The test consists of 36 photographs of people's eyes expressing different mental states. Alongside each photograph are four different mental state words (adjectives and nouns). The participant has to select the word they think best represents the mental state of the eyes shown in each photograph. The task is scored with 1 point for a correct answer and 0 points for an incorrect answer. The maximum score is, therefore, 36 points. Internal reliability was high for this task (Cronbach's $\alpha = .89$).

Figure 2. 2

Example Reading the Mind in the Eyes Test Stimulus



Mental state reasoning (Strange Stories)

A short-form version of Strange Stories (White et al., 2009) was used as a measure of mental state reasoning (see Figure 2.3). The eight vignettes used in the current study required an understanding of white lies, double bluffs, persuasion, and misunderstanding. After reading each vignette silently, participants were asked a single mental state question about why a character acted a certain way or said something that is not literally true. Responses were typed using a computer. Response accuracy was rated on a 0–2 scale by two researchers: 0 points for an incorrect answer, 1 point for a partially correct answer, and 2 points for a fully correct answer. The maximum score is therefore 16 points. Internal reliability for this task in the current study was low (Cronbach's $\alpha = .52$).

Figure 2. 3

A Vignette from the Strange Stories

Instructions: 'Here are some stories, and some questions. Read the stories silently, and I'll ask you some questions at the end of each story.' (Happé, 2009)

Simon is a big liar. Simon's brother Jim knows this. He knows that Simon never tells the truth! Now yesterday Simon stole Jim's ping-pong paddle, and Jim knows Simon has hidden it somewhere, though he can't find it. He's very cross. So he finds Simon and he says, 'Where is my ping-pong paddle? You must have hidden it either in the cupboard or under your bed, because I've looked everywhere else. Where is it, in the cupboard or under your bed?' Simon tells him the paddle is under his bed.

Q: Why will Jim look in the cupboard for the paddle?

Interpreting social situations in a narrative text (Short Story Task)

The Short Story Task mental state subsection (Dodell-Feder et al., 2013) aims to assess individual differences in reasoning about a range of mental states that differ in complexity during social interactions within a complex fictional narrative. In this task, participants read Ernest Hemingway's literary text 'The End of Something' (1,427 words), and they then answer questions about the story. The participant can refer back to the story at any time. There were two sections: a) the spontaneous mental state inference section and the mentalising section. Participants were asked a single question in the spontaneous mental state inference section, this was to summarise the story they had just read. Their answers were coded for mental state inferences; they scored 1 point if the

participant's description contained a mental state inference and zero if they did not. The mentalising section contains eight items with scores assigned based on the accuracy of the mental state inference. Each item is scored from 0–2 points (maximum possible score: 16). Internal reliability in this study was low (Cronbach's $\alpha = .54$). When compared to more established tests of affective mentalising, namely the Reading the Mind in the Eyes Test (Baron-Cohen et al., 2001) and the Interpersonal Reactivity Index (fantasy subsection; Davis, 1980), the Short Story Task has shown good concurrent validity in previous studies (Dodell-Feder et al., 2013; Dörrenberg, et al., 2018). A web-adapted version of the Short Story Task was created. Unlike the original version, responses were written in a text box rather than stated verbally.

Reading comprehension measures

Narrative text comprehension (Short Story Task)

The Short Story Task reading comprehension subsection (Dodell-Feder et al., 2013) measures general comprehension without reference to characters' mental states. The five reading comprehension items probe the participants' understanding of non-mental state story content. 0 points are given for inaccurate responses, 1 point is given for partial understanding, and 2 points are given for full understanding (maximum score 10). Internal reliability for this subscale in the current study was very low (Cronbach's $\alpha = .31$). A web-adapted version of the Short Story Task was created. Unlike the original version, responses were written in a text box rather than stated verbally.

Expository text comprehension (Adult Reading Test-2)

All participants were administered the silent reading version of the Adult Reading Test-2 (ART-2; Everatt, et al., 2017). In this version of ART-2, participants are asked to read an expository text silently and then answer questions about it afterwards. The test starts with one practice story (59 words long), which is followed by four question items. This is followed by three experimental stories (Test 1 is 179 words long, Test 2 is 268 words long, and Test 3 is 506 words long). Ten question items then follow. In each of the three sections, the participant reads the story silently, and the time it takes for them to read is recorded. The speed of reading scores can be calculated for each test story. The question items contain factual questions, memory questions, and inference questions. The ART-2 test manual reports reliability coefficients of 0.86 for comprehension and 0.90 for reading rate. The ART-2 also reports good convergent and concurrent validity (Brooks et al., 2017). The silent version of the ART-2 takes approximately 15 minutes to complete.

In the web-adapted version for the current study, the practice section and the first two sections were used, but the third section was removed to limit time. Once read, the story was removed and could not be referred to. The participant saw each question one at a time and can write their answer in a text box. As only two sections are used, the maximum score was 20. Internal reliability in the current study was good (Cronbach's $\alpha = .70$).

Control measures of language comprehension

Vocabulary knowledge (LexTALE)

All participants were administered the Lexical Test for Advanced Learners of English (LexTALE; Lemhöfer & Broersma, 2011). This is a computerised lexical decision test. In this test, participants saw a word on the screen and had to decide if it is a word from the English language (e.g., generic) or a non-word (e.g., platery). The test comprises 60 trials (40 words and 20 non-words). When processing the LexTALE results, the scores of the 40 words and 20 non-words are turned into percentages, and the total of this score is divided by 2. Therefore, a maximum score of 100% correct is possible. This task showed good internal reliability (Cronbach's $\alpha = .82$).

Inference generation (Implicit Management Task)

The original French version of *La gestion de l'implicite* (Implicit Management Task; Duchene, 2000) was translated into English for the purposes of the current study by a bilingual French-English speaker (see Figure 2.4). This task was used as a measure of inference generation. The task consists of 20 short texts; each text comprises a short narrative involving a scene with two speakers or a description of verbal interaction. These narratives contain affirmative statements which define a fact or present a problematic situation. The participants listened to an audio recording of each text, after which they heard three statements that related to the passage. Participants had to state whether these statements were 'True', 'False', or 'We don't know' in the case that the statement could not be answered from the information in the text. The statements were divided into three categories: inference, explicit (i.e. literal comprehension) and distractor. To answer inference questions correctly, the participant had to integrate

information provided in the text with background knowledge; the explicit statements required the participants to remember information directly from the text, and the distractor questions required the participants to identify that the statements could not be proven or disproven from information in the text. There are 60 statements in total, 26 of which are inference questions, giving a possible score of 26. Internal reliability was good (Cronbach’s $\alpha = .73$).

Figure 2. 4

Example Passage from the Implicit Management Task

Jack walked into the room in his pyjamas. He was very happy to see so many presents for him under the huge Christmas tree, which he decorated the night before with his parents.

a. Jack is outside. True False

WDK

b. Jack’s pyjamas are blue. True False

WDK

c. It’s the first day of spring. True False

WDK

Procedure

All testing took place online, and the battery took approximately 85 minutes to complete. All participants completed all tasks in the same order. The participant started the experiment by first reading the participant information sheet and then giving consent.

After this, they filled out a questionnaire that asked for demographic details (e.g., ethnicity, gender, occupation). The tasks were then completed in the following order: the Short Story Task, the Director's Task, the Implicit Management Task, ART-2, the Reading the Mind in the Eyes Test, Strange Stories, and finally LexTALE. There were also three attention checks (IMCs).

Interrater reliability

Two coders coded the Short Story Task. In the initial session, data from 25 participants were coded. Discrepancies were then discussed and resolved in a meeting between the coders, who then coded the remaining 69 participants' data independently. In this second phase of coding, a high degree of reliability was found between the two coders using intraclass correlations. The average measure was .92, with a 95% confidence interval of .87 to .95 ($F[68, 68] = 12.26, p < .001$). Any remaining discrepancies were then discussed and agreed upon by the two coders.

The two coders also coded Strange Stories. During the first phase of coding, each coder marked 25 participants. A high degree of reliability was found between the two coders in the intraclass correlations. The average measure was .93, with a 95% confidence interval of .93 to .99 ($F[1, 24] = 32.79, p < .001$). During the second phase of coding, the remaining 69 participants were coded, and a high degree of reliability was found between the two coders. The average measure was .98, with a 95% confidence interval of .97 to .99 ($F[1, 68] = 51.90, p < .001$). Any discrepancies were then discussed and agreed upon by the two coders.

2.4 Results

All results in Study 1 were analysed in SPSS, version 27 (2020). Descriptive statistics for all measures are presented in Table 2.2.

Table 2. 2

Descriptive Statistics for all Measures and their Subcomponents

Measure	Mean (SD)	Maximum possible score	Range
DT Accuracy Percentage Difference	18.09 (28.38)	N/A*	N/A*
Reading the Mind in the Eyes Total	25.71 (4.80)	36	23.31
Strange Stories	12.61 (2.33)	16	10.51
Short Story Task Mental State	9.69 (2.53)	15	12.93
Short Story Task Comprehension	8.62 (1.39)	10	5.71
ART-2 Total Score	14.50 (3.08)	20	15.07
LexTALE	88.33 (9.70)	100	42.05
IMT Inference	20.21 (3.36)	26	6.17

Note. DT = Director's Task, Short Story Mental State = Short Story Mental State Reasoning Subsection, ART-2 = Adult Reading Test, 2nd Edition, IMT = Inference-Making Test (Implicit Management Task).

As participants were recruited from two different populations (SONA and Prolific), the results of each group were first analysed separately. The results indicated that the mean scores on measures associated with vocabulary and reading

comprehension were significantly higher in the Prolific Group than in the SONA Group (see Table 2.3).

Table 2.3

Comparisons Between SONA and Prolific Participants on Measures Related to Reading Comprehension

	Platform	Mean (SD)	Max Score	T-Score
Short Story Task Comprehension	SONA ^a	8.28 (1.59)	10	$t(76.20) = -2.23, p = .029$
	Prolific ^b	8.92 (1.12)		
ART-2 Total	SONA ^a	13.41 (3.22)	20	$t(83.03) = -3.37, p = .001$
	Prolific ^b	15.46 (2.62)		

a (n = 44); b (n = 50)

Participants recruited from Prolific were older ($M = 40.18$, $SD = 12.76$) than the student participants recruited from SONA ($M = 19.84$, $SD = 2.95$). As such, group differences were further investigated by examining correlations between age and the measures of reading comprehension and mentalising. When examining the correlations between age and measures related to reading in each group (Prolific vs SONA) we can see that in the Prolific group there were significant correlations between age and (1) vocabulary, (2) Reading the Mind in the Eyes Test, and (3) the Short Story Task, reading comprehension subsection (Tables 2.4 and 2.5). These differences in performance on some tasks between the two samples are likely because the Prolific participants were older and had better vocabulary knowledge (See Discussion Section 2.5).

Table 2. 4*Sona - Correlations Between Age & Measures Related to Reading Comprehension*

	Director's Task	Strange Stories	RMET	Short Story Task comprehension	ART-2 Total	LexTALE
Age	.12	.07	.21	.29	.12	.16

p* < .05; *p* < .01

Note: RMET = Reading the Mind in the Eyes**Table 2. 5***Prolific - Correlations Between Age & Measures Related to Reading Comprehension*

	Director's Task	Strange Stories	RMET	Short Story Task comprehension	ART-2 Total	LexTALE
Age	-.12	.15	.31*	.28*	.13	.38**

p* < .05; *p* < .01

Note: RMET = Reading the Mind in the Eyes

Table 2. 6

SONA and Prolific groups - Correlations Between Mentalising, Reading Comprehension, and Control Variables

	1.	2.	3.	4.	5.	6.	7.	8.
1. DT		.28	-.23	.11	.24	.18	-.03	.44**
2. Reading the Mind in the Eyes	-.15		.25	.17	.31*	.10	.35*	.19
3. Strange Stories	.23	.24		.04	.30	.31*	.27	.07
4. Short Story Task – mental state	.12	.12	.24		.34*	.19	.08	.04
5. Short Story Task – reading comp.	.08	.36*	.23	.09		.31*	.15	.27
6. ART 2 – total score	.25	.10	.51*	.18	.14		-.32	.35*
7. LexTALE	-.06	.65**	.20	.29*	.35	.04		-.04
8. IMT	.07	.13	.34*	-.17	.17	.28	.03	

* $p < .05$, ** $p < .001$, $N = 44$ in SONA group, $N = 50$ in Prolific group

Note: RMET = The Prolific group correlations are in black text, whereas the SONA correlations are in blue text.

The correlations between all the variables in the Sona and Prolific populations can be seen in Table 2.6. In both groups, we can see small but non-significant correlations between the four mentalising variables. Additionally, when looking at the correlations between the mentalising variables and the language variables, significant correlations between the Short Story Task mental state subsection and the Short Story Task reading comprehension subsection (.34, $p < .05$) were found in the SONA group but not the Prolific group. Additionally, there was a small significant correlation between the Short Story mental state subsection and LexTALE (.29, $p < .05$) in the Prolific group but not the SONA group. There was also a significant correlation between

the Director's Task and the Implicit Management Task (inferencing) (.44, $p < .001$) in the SONA group but not the Prolific group. In conclusion, although there were some differences between the correlations in the SONA and Prolific groups, most of the correlations followed a similar pattern. For example, in both groups, there were only small correlations between the mentalising variables, which were not significant in either group. The correlations were similar between the mentalising variables and the reading comprehension measures. There were small significant correlations between the Reading the Mind in the Eyes Test and the Short Story reading comprehension section in both groups. Additionally, there was a small significant correlation between the Strange Stories and the ART 2 in the SONA group (.31*) and a moderate significant correlation between the two variables in the Prolific group (.51*).

Furthermore, the SONA group consisted of females only, whereas the Prolific group was mixed (22 females, 28 males). As research shows that females are more voracious readers than males (Bavishi et al., 2016), a series of independent t-tests were run to explore whether there were group differences between females and males in reading comprehension (ART 2 and Short Story reading comprehension subsection), vocabulary (LexTALE), and inferencing (Implicit Management Task). The t-tests found no significant differences between females and males in reading comprehension, $t(92) = -.22, p = .83$, vocabulary $t(92) = 1.00, p = .32$, or inferencing, $t(91) = .93, p = .35$. As the correlations between the variables were roughly similar, and there were no significant gender-based differences between SONA and Prolific groups, the decision was made to join the two populations in the regression analysis.

To explore the factor structure of the mentalising variables, a factor analysis was performed. As correlations were found between two of the mentalising variables, principal component analysis with Oblimin extraction was utilised. The resulting model

produced two factors explaining 60% of the variance (Kaiser-Meyer-Olkin = .54; $\chi^2 [6] = 9.63, p = .141$). This model was not significant, which suggests there was little correlation between the mentalising variables. On this basis, the four measures of mentalising were retained as separate variables for subsequent analysis since there was no rationale for computing a composite variable given their minimal coherence.

To explore the factor structure of the Short Story Task and the spontaneous, mentalising, and reading comprehension subsections, a factor analysis was performed. As items on the task were theoretically correlated, principal component analysis with Oblimin extraction was utilised. The resulting model produced six factors explaining 63% of the variance (Kaiser-Meyer-Olkin = .580; $\chi^2 [91] = 148.92, p < .001$). Although the model was significant, the mental state items did not load onto a single factor, indicating that the items within each subsection measured different constructs.

Finally, to test whether mentalising predicted reading comprehension, two multiple regression analyses were performed. The first regression assessed whether mentalising predicted reading comprehension in narrative texts, with the Short Story Task reading comprehension subsection as the outcome variable. As age and platform were positively correlated with mentalising, reading comprehension, and vocabulary, these two variables were added in step one. In the second step, the language variables (inference generation and vocabulary) were added. Finally, in the third step, the measures of mentalising were added. For this analysis, the Short Stories mentalising subscale was not included as a predictor, because this variable was extracted from the same task as the outcome variable (i.e., the reading comprehension subscale of the Short Stories task). The regression model predicting narrative text comprehension is summarised in Table 2.7. The combined set of predictor variables explained 29% of the variance of the Short Story Task reading comprehension score ($F [7, 85] = 5.02, p <$

.001; $R^2 = .293$). The mentalising variables accounted for a unique 7% of the variance in reading comprehension of narrative texts. No individual predictor variable reached significance.

Table 2. 7

Regression Model Predicts Variation in Reading Comprehension Using Narrative Texts

		Model 1: Outcome = Short Story Task Reading Comprehension				
	Predictors	B (SE)	β	t	ΔR^2	F-Change
Step 1	Age	.03 (.01)	.31	2.28	.111	.005
	Platform	.08 (.38)	.03	.20		
Step 2	LexTALE	.01 (.01)	.05	.48	.219	.004
	IMT	.14 (.04)	.33	3.43		
Step 3	DT	.01 (.01)	.01	1.42	.293	.037
	RMET	.06 (.03)	.03	1.96		
	Strange Stories	.09 (.06)	.16	1.63		

* $p < .05$; ** $p < .01$

Note. IMT = Implicit Management Task, DT = Director's Task, RMET = Reading the Mind in the Eyes Test

The second regression model assessed whether mentalising predicted reading comprehension of expository texts, with the ART-2 as the outcome variable (summarised in Table 2.8). Again, age and platform were added in step one. In the second step, the inference generation and vocabulary measures were added. Finally, the measures of mentalising were added in the third step, this time including the Short Story Task mental state subscale.

Table 2. 8*Regression Model Predicting Variation in Reading Comprehension of Expository Texts*

		Model 1: Outcome = ART-2 Reading Comprehension				
	Predictors	B (SE)	β	<i>t</i>	ΔR^2	<i>F</i> -Change
Step 1	Age	.03 (.30)	.15	1.07	.123	.003
	Platform	1.43 (.84)	.23	1.71		
Step 2	LexTALE	-.02 (.03)	-.06	-.56	.056	.054
	IMT	.22 (.09)	.24	2.44		
Step 3	DT	.02 (.01)	.17	1.85	.166	<.001
	RMET	-.01 (.07)	-.01	-.07		
	Strange Stories	.47 (.12)	.36	3.81*		
	Short Story Mental State	.14 (.11)	.11	1.22		

* $p < .05$; ** $p < .01$

Note. IMT = Implicit Management Task, DT = Director's Task, RMET = Reading the Mind in the Eyes Test.

The combined set of predictor variables explained 35% of the variance of the ART-2 score ($F [8, 85] = 5.62, p = <.001; R^2 = .346$). The mentalising variables accounted for 16% of the variance in reading comprehension of expository texts. Strange Stories was the only individual variable that reached statistical significance.

2.5 Discussion

This study had two aims: (1) to investigate the coherence of several different measures of mentalising and (2) to investigate the predictive role of mentalising in reading comprehension in an adult population. To fulfil these aims, the associations between four different aspects of mentalising (visuospatial perspective-taking, attributions from non-verbal cues, mental state reasoning, and interpreting social situations within an extended narrative text) and reading comprehension were investigated, while controlling for salient aspects of language comprehension (i.e. vocabulary knowledge and inference generation). It was expected that mentalising would be more strongly associated with the reading comprehension of narrative texts than expository texts.

Aim 1: Evaluating coherence between four mentalising tests in an adult sample

In line with findings reported by Warnell and Redcay (2019), minimal coherence was identified between the different mentalising tasks in adults; this was reflected in the small, non-significant correlations observed between the mentalising tasks and the results of a factor analysis indicating that the mentalising variables did not reliably load onto a single factor. This finding suggests that these different mentalising tasks test different aspects of mentalising and associated abilities. The results, therefore support the conceptualisation of mentalising as a multidimensional construct sustained by other cognitive abilities rather than as a unitary construct.

Participants' scores on two widely used mentalising tasks, Strange Stories and Reading the Mind in the Eyes, were weakly but significantly positively correlated, which is in line with previous research (Kirkland et al., 2012), implying some overlap in the

skills tapped by these tests. In both tasks, the participant must infer complex cognitive and emotional states from text-based or image-based stimuli and then describe them. Notably, both tasks also showed positive correlations with vocabulary and inference generation, suggesting that language skills may contribute to performance on both tasks. Previous research also indicates that the Reading the Mind in the Eyes Test is associated with vocabulary knowledge (Peñuelas-Calvo et al., 2019), which may further explain why it is more robustly associated with the text-based mentalising task (Strange Stories) than the test of visuospatial perspective taking (the Director's Task). The Director's Task did not correlate with Strange Stories or Reading the Mind in the Eyes Test and showed only small, non-significant correlations with the Short Story Task mental state subsection. This is perhaps not surprising given that the Director's Task does not require the participant to make inferences about cognitive and emotional states. It is likely that visuospatial mentalising is a separate component of mentalising. In addition, the Director's Task did not correlate with vocabulary knowledge, which seems reasonable as only simple instructions (e.g., 'move the ball up') are used in the task, but it did correlate with inference generation. This could potentially indicate that inferring the meaning of visuospatial instructions contributes to performance on this task.

The results indicate that the Short Story Task mental states subscale was not a reliable or valid measure of mentalising. This variable did not correlate with the other three measures of mentalising, including the Strange Stories which, at face value, has similar task demands (i.e., reasoning about mental states from passages of connected text). Further, a factor analysis determined that there was little correlation between the individual items on the Short Stories mental states subscale, and internal reliability was low. Unlike the other mentalising tasks included in this study, the Short Story Task did not correlate with the measure of inference generation. In the regression analysis, the

Short Story Task mental state subscale was not found to be a unique predictor of reading comprehension when controlling for vocabulary and inferencing abilities. Overall, these findings indicate that this task is not a reliable test of mentalising. The format and coding system of the Short Story mental state subscale hint that it may be better described as a test of reading comprehension (Turner & Felisberti, 2017), although its association with a standardised measure of reading comprehension (ART-2) was also low. The Short Stories task has been used in studies investigating mentalising in adults (Doddell-Feder et al., 2013), but its psychometric properties indicate that its utility in measuring individual differences in advanced mentalising is questionable.

Strange Stories is an established mentalising test that has been used in clinical, childhood, and adult studies (Wang & Su, 2006; White et al., 2009; Wilson et al., 2018). It has a more uniform format and coding system than the Short Story Task Mental State subscale. The text stimuli in Strange Stories are intentionally structured to elicit understanding of a specific mental state phenomenon (e.g., double bluff). The task comprises multiple vignettes, each followed by a single item in which the participant must explain why a character says something that is not literally true. In contrast, the narrative story used in the Short Story Task was not designed as a psychometric measure of mentalising, but is rather an excerpt from a single literary novel. Furthermore, as previous researchers have argued, the coding system does not distinguish between different types of mentalising (e.g., first- and second-order attributions or cognitive or affective mentalising) (Turner & Felisberti, 2017). If the Short Story Task mental state subscale were to use multiple numbers of shorter texts and specify a more precise coding system for responses, this might improve its validity. Overall, both measures of mental state reasoning from text showed poor reliability in the current study; the validity of the Short Stories task as a test of mentalising is also questionable. On balance, given its

extensive use in previous adult studies, the Strange Stories task is preferred for the remaining studies in this thesis, despite its psychometric weaknesses.

In summary, the four measures of mentalising selected in this study showed minimal coherence; there was no evidence that they are indicators of a unitary construct.

Aim 2: Examining mentalising as a predictor of reading comprehension in adults

The second aim of this study was to investigate the predictive role of mentalising in reading comprehension in an adult sample. It was predicted that mentalising would be a stronger predictor of comprehension of narrative texts than expository texts, in line with the framework outlined by Dore et al. (2018). This hypothesis was not supported. After controlling for age and language skills, it was found that one index of mentalising, mental state reasoning (measured by the Strange Stories), predicted comprehension of expository texts but not narrative texts. Specifically, the hierarchical regression showed that the set of predictor variables (demographic factors, vocabulary, inferencing, and the four mentalising predictors) explained 35% of the variance in ART-2 scores. The third step of this hierarchical regression added four mentalising predictors (Director's Task, Reading the Mind in the Eyes, Strange Stories, and Short Story Task), and explained an additional 16% of the variance in reading comprehension. The Strange Stories was the only mentalising task that was a significant independent predictor, indicating that mental state reasoning may be related to reading comprehension of narrative texts in adults. However, none of the three mentalising predictors predicted comprehension of narrative texts. Here, the set of predictor variables explained 29% of the variance in the Short Story Task, reading comprehension subscale. Step 3 of the regression contained the three mentalising predictors (Director's Task, Reading the Mind in the Eyes, Strange Stories).

Model 3 (containing the three mentalising variables) explained an additional 7% of the variance in reading comprehension of narrative texts. However, no individual predictor variable reached significance.

The results of this study indicate that a specific subcomponent of mentalising, mental state reasoning, may be implicated in individual differences in reading comprehension in adults. These results build upon childhood research that suggests that mentalising supports reading comprehension in young childhood (Atkinson et al., 2017; Kim, 2017). This study also coincides with Kim's (2017) finding that mentalising predicts reading comprehension after controlling for inference generation, suggesting that mental state reasoning may have a special relationship with reading comprehension across the lifespan.

Limitations

One limitation of this study is that it only utilised four of mentalising tasks that draw on different aspects of mentalising. Therefore, it is possible that other subcomponents of mentalising not tested in this study also show associations with mentalising. For example, Atkinson et al. (2017) found that false belief understanding predicted reading comprehension ability concurrently and longitudinally in young children. It is possible that false belief understanding continues to support reading comprehension in adults. One barrier is the lack of mentalising tasks available which are specifically designed for adults and distinguish between different subcomponents mentalising. For example, the Faux Pas Task has assessed false beliefs and has been used in adult populations (Bottiroli et al., 2016). However, this task also measures violations of social norms. Therefore, if a relationship between the Faux Pas Task and reading comprehension does exist is unclear whether this is due to the violation of social

norms or false belief elements of the task. In future research, more testing is needed to explore the relationship between different subcomponents of mentalising and reading comprehension

Contrary to the predictions and findings in childhood research (Boerma, 2017; Kim, 2021), mentalising was more strongly associated with scores of expository text comprehension than narrative text comprehension. This unexpected finding may be explained by differences in reliability and validity between the standardised reading comprehension measure using expository texts (ART-2) and the non-standardised measure of reading comprehension using narrative texts (Short Story Task reading comprehension subscale). The Short Story Task, reading comprehension subscale, only contained one story followed by five items and was shown to have low internal validity (.31), which may indicate there needed to be more items on the task.

In contrast, ART-2 has three sections (only two sections were used in this study), each containing one passage and 10 items (balanced between memory, factual, and inference items), and in each section the passages increase in length and complexity. In addition, the internal consistency of the ART-2 was higher (.70), indicating it is a more reliable measure. These differences in the tasks' structures may mean that the Short Story Task reading comprehension subscale is less sensitive to detecting individual differences in reading ability. Additionally, ceiling effects emerged in the Short Story Task Reading Comprehension subscale but not in the ART-2. Hence, until a standardised measure of reading comprehension using narrative texts is designed for adults in the United Kingdom, it may be necessary to use different methods, such as measuring narrative recall of mental state terms (see Kim et al., 2021), to investigate the influence of mentalising in reading comprehension using different text types.

The results also indicated that there were differences in abilities between the subsamples recruited via SONA and Prolific, which likely arose from age. The SONA participants were recruited from a university and were younger (M age=19.82) than the Prolific participants (M age=40.18), who were recruited from the general population. Specifically, there were significant positive correlations between; age and vocabulary (LexTALE) (.379**), age and attributions from non-verbal cues (Reading the Mind in the Eyes Test) (.312*), age and reading comprehension using narrative texts (Short Story Task, reading comprehension subsection) (.280*). These correlations indicate that in the Prolific group, the older participants had better vocabulary, were better at comprehending the Hemmingway short story in the Short Story Task and were more accurate at selecting the appropriate mental state adjectives to describe the pictures of eyes in the Reading the Mind in the Eyes Test. Although findings are mixed, some research shows that mentalising in comparison to other cognitive (e.g., executive functions) is relatively preserved as we age, so this may explain why the Prolific participants who were older on average score better on the Reading the Mind in the Eyes Test (Giovagnoli, 2019; Happé, 1998).

Another reason the older participants scored more highly on the Short Story Task, reading comprehension subsection, and the Reading the Mind in the Eyes Test is likely due to the fact they have a broader vocabulary knowledge than the younger participants. Research shows that older adults have a wider vocabulary knowledge than younger adults (Laumann et al., 2000; Sternberg, 1987). Sternberg (1987) suggests that most vocabulary is learnt in context rather than studied, and as people get older, they have more time to experience new words in different contexts (e.g., reading for leisure or via the workplace) and to learn a larger vocabulary. Vocabulary is a foundational skill in reading (Kim, 2017). A wider vocabulary may have been particularly important when

reading the short story (*The End of Something*, 1925) as it is an older book, and the language may be unfamiliar to younger readers. Furthermore, the older participants' score on the Reading the Mind in the Eyes Test are likely related to their vocabulary. Vocabulary has been associated positively with scores on the Reading the Mind in the Eyes Test (see Section 3.4.6 for more detail) (Peterson & Miller, 2012; Olderbak et al., 2015). The relationship between vocabulary and the Reading the Mind in the Eyes Test is likely due to the non-mentalising elements of the task. Specifically, the participant must pick between four non-mental state adjectives to describe the pictures of the Eyes and participants who have a larger vocabulary knowledge seem more accurate at describing the eyes appropriately. These results, therefore, indicate that older participants may have an advantage when completing the Reading the Mind in the Eyes Test and the Short Story Task, reading comprehension subsection, due to their broader vocabulary knowledge.

Current research indicates that despite a decline in cognitive functions, reading comprehension is relatively well preserved in older adults (over 65 years old; Artuso, & Belacchi, 2021; De Beni et al., 2007). The results from the study support previous research findings that attributions from non-verbal cues (Li et al., 2019) and mental state reasoning appear to be preserved in older adults (Giovagnoli, 2019; Henry et al., 2013). It is therefore possible that mentalising supports reading comprehension in older adults. Another possibility is that a factor not measured in this study but associated with mentalising also supports reading at a later age. In young children, the experience of reading fiction has been associated with improved mentalising (Mar et al., 2010) and higher scores on Strange Stories (Hamilton et al., 2020). Boerma et al. (2017) also discovered that mentalising, as measured by the Strange Stories, mediated the relationship between the home literacy environment and reading comprehension ability

in middle childhood. As reading experience has been shown to support older adults' reading comprehension (Payne et al., 2012; Stanovich et al., 1995), it is possible that reading experience is a factor associated with the relationship between mentalising and reading comprehension.

Conclusion

The first aim of this study was to examine the relationship between four different measures of mentalising in adults. The results lend support to the idea that there is minimal coherence between different mentalising tasks in adults (Apperly, 2012; Warnell & Redcay, 2019). Accordingly, mentalising is better understood as a multidimensional construct supported by an array of different cognitive abilities (e.g., inferencing generation, vocabulary). As there has been little research in this area, these findings have important implications. Firstly, they indicate that future research should use multiple measures of mentalising to establish whether psychological phenomena (e.g., reading comprehension; clinical diagnoses such as autism or psychopathy) are linked to mentalising or whether effects are task-specific. A highly cited example of this is a study by Kidd and Castano (2013) that links brief exposure to fiction texts with improvements to mentalising ability, using only the Reading the Mind in the Eyes Test as an indicator of mentalising. Attempts to replicate this study have been mixed (Kidd & Castano, 2019; Panero et al., 2016). Second, the results of the current study indicate that more research must be undertaken to establish which non-mentalising skills underpin performance in different mentalising tasks. For example, the findings suggest that performance on Strange Stories and Reading the Mind in the Eyes Test is supported by

vocabulary and inference generation. In contrast, performance on the director's task is supported by inference generation but not vocabulary in adults.

The findings also call into question the reliability and validity of the Short Stories task as a measure of mentalising. This measure was selected as there are very few measures of mentalising specifically designed for adults. The Short Story Task is the only measure that could be sourced that assesses mental state reasoning using a sophisticated literary text. Therefore, it was deemed a potentially useful task to measure advanced mentalising in adults. However, analyses yielded only a small correlation ($r = .12$) between this task and the performance of the more established Strange Stories task. While both purport to measure reasoning about mental states, there are numerous potential reasons for the lack of coherence between the tasks, as discussed above. However, it seems likely that psychometric issues underpin the lack of coherence between these two tasks. The Short Stories test has not been validated, is based on only one text, and seems to be less sensitive than other measures, and it therefore requires further psychometric testing and development. To establish the extent of overlap and difference between these two tasks, and to understand the structure and properties of mentalising measures more broadly, a more thorough review of the mentalising tasks used with adults is needed (see Chapter 3).

The second aim of this study was to investigate the predictive role of mentalising in reading comprehension in adults when controlling for language skills. A related aim was to assess whether mentalising was more strongly associated with the comprehension of narrative or expository texts. Mentalising has been shown to predict reading comprehension in children (Kim, 2017), and there is emerging evidence that mentalising is associated with the comprehension of narrative texts specifically (Kim, 2021). In line with this, the findings indicate that mentalising skills, specifically mental state reasoning

(Strange Stories), are implicated in reading comprehension beyond childhood. Although mentalising was shown to predict performance in reading comprehension in both narrative and expository texts, the regression analyses indicate that mentalising has a greater role in the comprehension of expository texts. However, this may well be due to methodological differences between the standardised measure of reading comprehension using expository texts (ART-2) and the non-standardised measure of reading comprehension using narrative texts (Short Story reading comprehension subscale). For instance, high ceiling effects were found in the Short Story measure, and its internal reliability was low, indicating it was not a suitable measure of reading comprehension using narrative texts. Therefore, we cannot say with certainty from these results that mentalising does not have a role in reading comprehension of narrative texts. Sourcing reliable and valid measures of both advanced mentalising and (narrative) reading comprehension that are suitable for use with adults is challenging. The following chapter presents a systematic taxonomy of advanced mentalising measures in order to address some of these issues and to inform the design of future empirical studies of mentalising in adults.

3. Properties of mentalising tasks used with adults: A systematic review and taxonomy

3.1 Overview of the chapter

This chapter reports a systematic review of the tasks that have been used to measure individual differences in mentalising in adults. This review is needed because there is uncertainty about what different mentalising tasks measure. Indeed, recent research shows little coherence between different mentalising tasks, meaning that performance on one mentalising task does not predict performance on another. This suggests that the abilities required to perform well on mentalising tasks may differ depending on the task's properties (Bialecka-Pikul et al., 2021; Warnell & Redcay, 2019). Different tasks may index different subcomponents of mentalising (e.g., mental state reasoning; mental state attributions from non-verbal cues). The existing literature indicates two reasons why different mentalising tasks do not cohere. First, they measure different subcomponents of mentalising, such as taking a visuospatial perspective or interpreting social situations/ Second, because mentalising tasks use stimuli in various

formats (e.g., verbal stories, photographs, animated shapes), it is likely that successful performance recruits different non-mentalising skills (e.g., reading ability visual processing). Therefore, the lack of coherence among mentalising tasks may be explained by differences in the sub-skills drawn upon to successfully complete each task: mentalising subcomponents and/or non-mentalising abilities.

This systematic review aimed to collate mentalising tasks that have been used in research with adults over the last 20 years and to examine their properties. Assessing the subcomponents of mentalising that each task measures was key. A secondary aim of this review was to determine which mentalising tasks have been used to study the relationship between mentalising and language. Language is important, because many of the tasks used to measure mentalising involve listening to or reading stories; tasks that purport to measure mentalising and tests of oral or written language thus often have shared verbal task demands. Previous studies that have reported relationships between mentalising and other cognitive abilities, including language, may therefore be misleading, as it is unclear whether “mentalising” correlates with language or whether verbal task demands explain reported associations (Boerma et al., 2017; De Villiers, 2007; Ebert, 2020). Therefore, this systematic review aimed, to examine reported associations between different tasks of mentalising and language in detail, in order to better understand the nature of this relationship in adult samples. By (a) describing the subcomponents of mentalising tapped by different tasks and (b) examining correlations between different subcomponents of mentalising and language, the final aim of this review was to produce a taxonomy of tests of adult mentalising for use by researchers in designing studies of mentalising in adults.

3.2 Introduction: Development of mentalising after five years old

Over the past 20 years, research examining mentalising beyond the age of five years old has grown steadily and has challenged previous conceptions about the nature of mentalising and the underlying mechanisms. The construct of mentalising was originally understood from a conceptual perspective as qualitative stages of development that were observed as children reached critical milestones throughout childhood (e.g., understanding desires and false beliefs) (Astington & Jenkins, 1999; Wellman & Bartsch, 1988). These qualitative stages of development were observed universally in preschool children, with minimal variation observed in the ages at which these concepts were grasped. In addition, meta-analyses of neuroimaging studies with adults suggest that a uniquely human, independent mentalising network exists in the developed brain and controls representations of other people's mental states (Frith & Frith, 2006; Saxe, 2006). Together, these developmental and neuroscientific findings suggest that mentalising may exist as a unitary construct that develops through a coherent series of milestones across the lifespan.

However, the conceptual approach to mentalising was based on a body of neuroimaging research and developmental research that focused on a limited age range (two to five years old) and used a restricted number of behavioural measures (e.g., false belief tasks) (Apperly, 2012; Schaafsma et al., 2015). More recent neuroimaging research using a new range of techniques has found more distinct activation patterns, indicating that the mentalising network may be more fragmented than previously thought (Schaafsma et al., 2015). In addition, developmental researchers examining mentalising in older children and adults using a range of tasks found few conceptual milestones after the age of five years old, concluding that mentalising development was

more incremental and multidimensional from middle childhood onwards (Apperly, 2012; Gerrans & Stone, 2008; Warnell & Redcay, 2019). Therefore, to examine mentalising development after the age of five, researchers have adopted an individual differences approach to detect quantifiable differences in children's and adults' abilities to reason about mental states (Boerma et al., 2017; Milligan et al., 2007). To separate the relatively stable period of mentalising development in preschool children from the diverse development seen after this age, some researchers have used the term 'advanced mentalising' (Hughes & Devine, 2015).

More recent research has used a broad range of behavioural techniques to examine mentalising across the human lifespan (Osterhaus & Bosacki, 2022; Schaafsma et al., 2015; Schurz et al., 2014). Warnell and Redcay (2019) found that after the age of five, there is little sign of clear developmental pathways in subcomponents of mentalising or stages of universal conceptual understanding. Examining correlations among a range of mentalising tasks over three different age groups (preschool, middle childhood, and adulthood), the authors noted little coherence amongst the tasks; that is, performance on one task did not predict performance on another. Their results suggest that different mentalising tasks measure different abilities (see Section 2.2 for more detail). Warnell and Redcay (2019) argue that coherence amongst the different mentalising tasks may be lacking because they measure different sub-components of mentalising, such as visuospatial perspective taking and mental state reasoning. They also propose that these different subcomponents of mentalising may be associated with the development of different cognitive abilities, such as language (De Mulder et al., 2019), working memory (Lecce et al., 2017), and inhibition (Osterhaus et al., 2016), which may further explain the observed lack of coherence. Therefore, mentalising may not be best conceptualised as a single construct, but rather a multidimensional construct

comprised of a number of subcomponents. Each subcomponent (e.g., mental state reasoning; attributing mental states from facial expressions) coordinates different aspects of perspective-taking in interaction with different non-mentalising skills.

The lack of coherence among behavioural tasks of mentalising corresponds to new developments in neuroimaging research. The mentalising network was evidenced by meta-analyses investigating overlapping activation across studies (Frith & Frith, 2003; Goel et al., 1995). These studies found that three components of the brain were consistently activated during mentalising tasks: the medial prefrontal cortex, the temporal poles, and the posterior superior temporal sulcus. This suggested that all reasoning involving mental states was processed in one coherent network in the brain and supports the view of mentalising as a unidimensional construct. However, recent neuroimaging research has indicated that rather than one network in the brain regulating all mental state input, separate regions in the brain may be responsible for controlling specific subcomponents, or modalities, of mentalising, which would indicate that mentalising is a multidimensional construct. When Schaafsma et al. (2015) re-examined results from meta-analyses of neuroimaging studies more critically, they found high heterogeneity both in the mentalising tasks used to elicit brain activation and in the neuroimaging techniques used in these studies. Hence, earlier meta-analyses that found evidence of a mentalising system based on pooled neuroimaging may have missed activation patterns elicited by specific tasks. More recent meta-analyses have grouped studies based on the neuroimaging technique and mentalising tasks used (Schurz, 2014). These studies have revealed distinct activation patterns that were previously blurred in pooled data. For example, when participants were presented with the Reading the Mind in the Eyes Test (tapping mental state attributions from non-verbal cues), the left inferior frontal area was activated. In contrast, the bilateral temporoparietal cortices were

activated when participants were shown the Frith-Happé Animations Test (attributing intentions to moving shapes). If different regions of the brain are responsible for regulating different subcomponents of mentalising, rather than a single mentalising network regulating all reasoning about mental states, this would suggest that mentalising is a multidimensional construct.

3.2.1 How mentalising is measured

The way in which mentalising is operationalised in empirical studies varies widely (Wellman, 2018). Mentalising has been measured with tasks that primarily draw on pragmatic language comprehension (e.g., Strange Stories; White et al., 2009), interpreting social situations (e.g., The Short Story Task; Dodell-Feder et al., 2013), making mental state attributions from non-verbal cues (e.g., Reading the Mind in the Eyes, Baron-Cohen et al., 2001), and attributing intentions to moving shapes (e.g., Frith-Happé Animations; Abell, F., Happé, & Frith, 2000). Consequently, and perhaps unsurprisingly, coherence amongst these tasks is lacking, as there is likely wide variation in the mentalising and non-mentalising demands of each task. For instance, the Reading the Mind in the Eyes Test involves attributing mental states to pictures of peoples' eyes, which will recruit visual processing. By contrast, in the Strange Stories task, individuals must read vignettes describing a social situation and use characters' mental state information to interpret the intended meaning of a character's pragmatic language.

Recent developments in the way that mentalising is understood conceptually (e.g., as multidimensional rather than unidimensional) have not been reflected in research practice. The most frequently used mentalising measures over the last 10 years

(second-order false belief tasks, the Strange Stories task, the Hinting Task, the Reading the Mind in the Eyes Test, and the Frith-Happé Animations Test) were designed between the 1980s and early 2000s to measure mentalising as a unidimensional construct (Osterhaus & Bosacki, 2022). Furthermore, studies use tasks with different mentalising and non-mentalising properties interchangeably to measure mentalising. For example, the Strange Stories and Reading the Mind in the Eyes Tests clearly tap different component skills. Baron-Cohen et al. (2001) claim that the Reading the Mind in the Eyes Test was designed to measure the first stage of mentalising, i.e., attributing relevant thoughts and/or feelings to an individual (e.g., he looks angry). By contrast, the Strange Stories task measures the second stage of mentalising, inferring the reason for this emotion (e.g., he has just been fired). However, both tasks have been used to associate mentalising with reading fiction (Boerma et al., 2017; Kidd & Castano, 2013, 2019). Therefore, it remains unclear which aspects of mentalising are related to reading abilities. For a thorough exploration of the relationship between mentalising and reading abilities, researchers must specify which aspects of mentalising each task measures. Using multiple mentalising tasks in study designs would allow researchers to ascertain that associations with other skills (e.g., reading) are specific to mentalising, rather than the other shared non-mentalising skills (e.g., vocabulary knowledge).

To characterise the relationship between mentalising and reading comprehension accurately, therefore, it is necessary to break tasks down into their mentalising and non-mentalising constituent parts. This idea was proposed by Schaafsma et al. (2015), who stated that a detailed taxonomy of the basic components of mentalising tasks should be developed to gain a greater understanding of mentalising and its relationship with other cognitive skills. Such a taxonomy of behavioural tasks would also enable neuroscientists to map the neural correlates of mentalising more accurately.

3.2.2 Operational terms used in this systematic review

This systematic review was conducted considering recent research that (1) defines mentalising as the process of understanding the mental states and perspectives of others and (2) suggests that this process may be multidimensional and interact with other cognitive abilities, such as working memory (Apperly, 2012; Lecce et al., 2017; Warnell & Redcay, 2019). The labels assigned to each operational category (e.g., higher-order false belief reasoning and tasks that draw on pragmatic language) were devised to establish what core aspects or subcomponents of mentalising each task measured; this categorisation is similar to that used in Osterhaus and Bosacki's (2022) recent systematic review. Notably, individual tasks may often draw on several subcomponents of mentalising. For example, in the current review, the Strange Stories task has been labelled as a 'task that draws on pragmatic language'; however, it also elicits interpretation of the mental states of characters in social situations and, thus, could arguably be labelled as 'interpreting social situations' alongside the Short Story Task.

To determine the primary mentalising subcomponent that each task draws upon, consideration is given to the description of the coding scheme in papers describing tasks for the first time. For example, in their revision of the Strange Stories task, White et al. (2009) describe items as assessing understanding of why a character said something that was not literally true, which would suggest it tests pragmatic language abilities. This description still appears to be relevant. Baron-Cohen et al. (2001) describe the Reading the Mind in the Eyes Test as measuring a person's ability to attribute mental states (e.g., compassion) to another person, without reference to the context (e.g., compassion following the loss of a loved one). However, attributing a mental state to a person's eyes

may involve different skills (e.g., visual processing) than attributing a mental state to a person screaming from another room (e.g., auditory skills). Therefore, the Reading the Mind in the Eyes Test was operationalised as ‘attributions from non-verbal cues’.

Although producing distinct categories can be misleading, as individual mentalising tasks may draw upon multiple mentalising and non-mentalising abilities, this review attempted to create operational terms that reflect the current understanding of mentalising as a multidimensional construct. Table 3.1 presents the different ways in which mentalising concepts have been operationalised, along with examples of the tasks used to measure them and a description of the tasks’ coding systems.

Table 3. 1

Operationalisation of Mentalising

Operationalisation	Example Task	Description	Coding
Higher-order false belief reasoning	Higher-order false belief reasoning using vignettes	False belief tasks measure a person’s ability to understand that another person can hold a mistaken belief (Bibby & McDonald, 2005). First-order false belief understanding is usually acquired at four years, but the ability to understand more sophisticated second-order false belief reasoning (e.g., why does Mary think John thinks about...?) or higher-order false belief reasoning has been	False belief tasks are typically coded as binary pass/fail (1/0) (Baron-Cohen et al., 1985).

shown to develop in adulthood
(Wellman & Lui, 2004).

Tasks that draw on pragmatic language	Strange Stories	Tasks that draw on pragmatic language measure a participant's ability to interpret the intended meaning of an utterance, which may differ from the literal meaning. For example, in the Strange Stories task (White et al., 2009), participants read a vignette describing a social situation and are then asked to answer one question regarding why a character said something that is not literally true (e.g. sarcasm, white lie).	There are eight items in total and a maximum score of 16. In the Strange Stories task, participants are awarded 2 points for a correct answer that references the characters' mental state, 1 point for an answer that references facts in the story, and 0 points for an answer that references irrelevant or incorrect information.
Interpreting social situations	The Short Story Task	Tasks involving interpreting social situations usually present participants with a social scene wherein characters interact. This scene may take the form of a vignette or a film. For example, in the Short Story Task (Dodell-Feder et al., 2013), the participant reads a literary short story about a couple breaking up and answers questions regarding the individuals' mental states and emotions.	There are 14 items in total, with a maximum score of 27. There is one spontaneous mental state question, in which the participant describes the story (1 point if they mention a mental state, 0 points if not). There are five comprehension items (2 points for a full answer, 1 for a partially correct answer, and 0 for an incorrect answer). Finally, there are

eight mental state reasoning questions (2 points for accurately identifying the mental states of the characters, 1 point for partial understanding, and 0 points for an incorrect answer).

Attributions from non-verbal cues	Reading the Mind in the Eyes	Attributions from non-verbal cue tasks ask participants to attribute cognitive or affective mental states to visual stimuli , e.g., faces. For instance, in the Reading the Mind in the Eyes Test (Baron-Cohen et al., 2001), the participant is presented with a series of faces that display different emotions. The participant must select one of four adjectives that best describes the emotion portrayed in the facial cue.	This task contains 36 items, with a maximum score of 36. Participants are awarded 1 point for correctly identifying the adjective that matches the mental state expressed by the eyes and 0 points for an incorrect answer.
Visual perspective taking (involving humans)	The Director's Task	Referential communication tasks test a person's ability to be sensitive to a speaker's perspective during a spoken exchange of information. To successfully complete a referential communication task, the person must represent the speaker's visual perspective as well as their own	This task comprises 48 trials: 32 filler trials (distractors), eight control trials, and eight experimental trials. The participant can score a maximum of 48 points. The participant gains 1 point for correctly following the director's instructions (by

		(Leinonen & Letts, 1997). In the Director's Task, the participant is presented with a 4x4 set of shelves containing objects, some of which are only visible to the participant. The participant must then listen to instructions from the 'director' to move objects. The participant must only move objects that are visible to both themselves and the director (Dumontheil et al., 2010).	moving the correct object to the correct location) and 0 points for incorrectly following the director's instructions. The percentage-correct score is calculated by dividing the number of points scored in the experimental condition by the total score in the experimental condition (8 points).
Attributing intentions to inanimate objects	Frith-Happé Animations Test	These tasks are based on Heider and Simmel's (1944) finding that people tend to attribute narratives involving mental state interactions to inanimate objects, such as moving geometric shapes. In the Frith-Happé Animations Test (Campbell et al., 2006), the participant watches two triangles that act out scripted scenes. In the goal-directed animations, the triangles simply react to each other's physical movements (e.g., fighting or dancing), whereas in the mentalising animations, the triangles react to each other's	This task involves eight animations: four based on a mentalising interaction and four on goal-directed interactions. The maximum score on the task is 64. Responses are coded for appropriateness and the amount of intentional language used. First, the participant is given 1 point for correctly identifying the animation (mentalising or goal directed). Second, the

		mental states (e.g., surprising or seducing). The participant must describe what they see.	participant is scored regarding how appropriately they describe the scene (2 for correct, 1 for partially correct, and 0 for incorrect). Finally, the participant is awarded points for the intentional language they used (0–5 points) for each animation.
Recognising violations of social norms	Faux Pas Task	Recognising violations of social norm tasks require the participant to recognise that a social transgression has occurred. An example is a faux pas, which is when a person unconsciously offends another person by saying something they do not want to hear (e.g., mistakenly stating that someone is older than they are). In the Faux Pas Task (Baron-Cohen et al., 1999) the participant reads a vignette that describes a social interaction wherein a faux pas occurs. To respond correctly, they must recognise the comment that has caused offence and attribute the correct emotion to the offended individual. This requires knowledge of social norms.	There are 10 vignettes, each followed by four items that assess the following: (1) the participant's ability to recognise that a faux pas has occurred, (2) the participant's ability to identify what the faux pas was, (3) a reading comprehension question (control), and (4) a final question to check whether the participant understood that the faux pas was a consequence of the speaker's false belief rather than an action with malicious intent. The participant scores 1 point for a correct answer and 0 points for an incorrect

answer. The maximum score on the task is 40.

Use of mental state terms in expressive language	ToM Story	In a ToM narrative production task, the participant is asked to make up an imaginary story based on a social situation. Their responses are recorded and then coded based on the amount of mental state language they use (Andreasen et al., 2008).	Each participant's story is separated by sentence. Each sentence is then rated as having no mental state attributions (0 points), one mental state attribution (1 point), or more than one mental state attribution (2 points).
Self-evaluated mentalising	Zinch mentalising interview	Mentalising interviews and questionnaires consist of questions about the 'self' and 'other' (Zinck et al., 2021). Questionnaires consist of questions that measure the respondent's self-reported tendency to adopt the perspective of others.	The interview comprises four sections: (1) introductory questions, (2) questions relating to the 'self', (3) questions relating to a close 'other', and (4) closing questions. The participant's responses are divided into chunks and then coded as physical (interests and facts) or mentalistic (psychological, relationship, and meta-statement).

3.2.3 The relationship between mentalising and language

The sociocultural view of mentalising development suggests that language plays a fundamental role in the development of mentalising as we develop through childhood into adulthood. Heyes and Frith (2014) describe the acquisition of explicit mental state understanding as a slow, arduous process during which children learn about mental state concepts and causal relationships through conversations with peers and adults. Indeed, Vygotsky's concept of higher-order mental functions has been recruited to explain the development of mentalising, which suggests that the growth of cognitive structures is built through interpersonal communication and relationships. According to Fernyhough (2008), Vygotskian ideas of internalisation and the mediating role of semiotic systems (e.g., language) may be particularly important for the development of mentalising. Internalisation refers to the process through which an individual internalises shared dialogic experiences with others, which shapes their own knowledge of the world. Children will then retain these internalised dialogic experiences as inner speech which they will then abbreviate (Visser et al., 2020). For instance, a child will internalise abbreviated dialogic exchanges with their mother during a joint problem-solving task and then use this acquired knowledge when engaging in solo activities (Wertsch & Stone, 1985). During such solo activities, the child can be heard verbalising knowledge learnt from these joint exchanges (e.g., 'Where does this go?') as they attempt to problem solve. The use of self-directed speech to guide and self-regulate behaviour during solo activities is known as private speech and has been shown to correlate positively performance on problem solving tasks (e.g., Tower of Hanoi) (Fernyhough & Fradley, 2005). In addition, semiotic systems, such as language, may mediate and enhance the development of pre-existing cognitive structures that support mentalising, by allowing individuals to develop their understanding of social processes

during joint exchanges. Fernyhough (2008) suggests that through spoken dialogue with others, people can describe their perceptions of reality and understand the perspectives of others; hence, humans have a uniquely efficient way of understanding each other (Clark, 1998). Therefore, from a young age, language may play a key role in developing an individual's ability to represent and reason about others' perspectives.

In the dialogic thinking framework, Fernyhough (1996) draws on the work of Vygotsky to argue that higher-order processes, such as mentalising, are dialogical processes that have been acquired and developed through interpersonal interactions. According to this framework, inner dialogue develops by internalising dialogically based interactions between individuals. Fernyhough suggests that inner dialogue is acquired developmentally, first through social speech between individuals (e.g., mother–child), then through private speech, and finally through conversion of inner verbal thought. Fernyhough and Meins (2009) demonstrate that private speech correlates positively with mentalising at four years – the age at which internalised speech is typically acquired – but negatively at 6 years old, as use of private speech at this age would suggest a developmental delay in internalisation. From an evolutionary perspective, systematically framing thoughts in dialogic terms may have evolved as it helps individuals organise their internal representations efficiently and reduces processing costs (Clark, 1998). Fernyhough (2008) explains that individuals may use their ability to conduct internal dialogues to simultaneously represent multiple perspectives of others during social interactions. Crucially, this framework emphasises that discourse-based interactions of any kind promote the development of mentalising at a young age if perspectives are discussed.

It has been noted that the age at which children pass false belief tasks – approximately four years old – coincides with the age at which they are able to use mental state verbs (e.g., think, know, and believe) in sentences in a sophisticated manner (Astington

& Jenkins, 1995; de Villiers, 1995). On the one hand, it has been suggested that the proper use of mental state verbs is preceded by the understanding of these mentalising concepts, such as false beliefs (Tager-Flusberg, 1993). On the other hand, researchers have found that the acquisition of certain aspects of language is necessary for a child to grasp the concept of false beliefs. Peterson and Siegal (2000) propound that to pass false belief tasks, children must understand mental state terms (e.g., believes, intended). Others have suggested that children must be able to use mental state terms in syntactically correct sentences (De Villiers & De Villiers, 2000; Schick et al., 2007). Much of the research in this area has focused on the relationship between language and the age at which children gain a conceptual understanding of false beliefs, which is generally at four to five years old. However, beyond this age, there is little evidence of conceptual milestones of mentalising development, and it is thus harder to examine the correlation between language and conceptual developments. To characterise how the relationship between mentalising and language matures develops after early childhood, researchers have used an individual differences approach.

The individual differences approach posits that some people will be better at reasoning about mental states than others and that these differences can be quantified. Yet, a major obstacle in this field is the uncertainty regarding what mentalising tasks measure, as the properties of mentalising tasks are currently poorly described. As a result, previous studies that have shown relationships between mentalising and other cognitive abilities, such as language, may be misleading, as it is unclear which subcomponents of mentalising may correlate with language or whether the non-mentalising aspects of the task may explain the associations. For example, mentalising has been associated with spoken and written language abilities from a young age into adulthood (Boerma et al., 2017; Hochstein et al., 2018; Trott & Bergen, 2020). However, these studies all used text-based measures of mentalising, such as the Strange Stories task and the Short Story Task (Dodell-Feder et al.,

2013); hence, the correlations might be explained by the verbal properties of the tasks rather than any conceptual relationships between mentalising and language. Although studies do control for language, the most frequently used control measure is vocabulary, which is a lower-order language skill. Therefore, higher-order or pragmatic aspects of language may be responsible for the association between language and mentalising (Bosco et al., 2018a; Osterhaus & Bosacki, 2022).

Research demonstrates that the Strange Stories task captures individual differences in mentalising: for instance, autistic children who pass false belief tasks have shown impaired performance on Strange Stories (Happé, 1994). In typically developing children, performance on the Strange Stories appears to increase with age, and the task has also captured individual differences in adults (Happé et al., 1998; Hughes, 2011; O'Hare et al., 2009). However, the task appears to draw on a range of abilities. For instance, Bosco argues that the Strange Stories task measures a mixture of mentalising concepts (e.g., false beliefs, white lies, and double bluffs) and pragmatic language concepts (e.g., irony and persuasion). As the vignettes are either read or heard aloud, the task also covers either reading or oral language comprehension. Other research suggests that the Strange Stories task draws on executive functions, specifically verbal fluency and deductive reasoning (Ahmed & Miller, 2011). However, the fact that the Strange Stories task draws on multiple abilities does not invalidate the task as a measure of mentalising, as it clearly measures an individual's ability to reason about the mental states of the characters in each vignette. It merely reflects the current theoretical understanding of mentalising as a multidimensional construct that is supported by a range of cognitive abilities.

To characterise the relationship between mentalising and language from an individual differences approach, researchers should not attempt to control for every skill that a mentalising task such as the Strange Stories task draws upon, as the assumption in doing

so is that mentalising is an ability that works in isolation (Apperly, 2012). Instead, researchers should attempt to determine and describe the skills that support mental state reasoning in each mentalising task to accurately portray how each of the many subcomponents of mentalising (e.g., taking a visual perspective and recognising violations of social norms) work in cooperation with other cognitive abilities. This would lead to an accurate depiction of the relationship between mentalising and language.

3.2.4 Aims of this systematic review

This systematic review aimed to explore which mentalising tasks have been used in research with adults and to examine the characteristics of these tasks in detail. This review built upon an earlier systematic review by Osterhaus and Bosacki (2022) that examined which mentalising tasks have been used most frequently over the last 10 years in research examining mentalising in children over the age of five and adults. A key finding from their review was that mentalising has been operationalised in a heterogenous manner.

Additionally, they found that a small number of established mentalising tasks (false belief tasks, the Strange Stories task, the Hinting Task, the Reading the Mind in the Eyes Test, and the Frith-Happé Animations Test) accounted for 60% of the mentalising measures used across all the studies selected in their review. Moreover, these tasks were used interchangeably to measure a single construct (mentalising), which is problematic, as recent research highlights that mentalising is a multidimensional construct. Therefore, mentalising tasks may draw upon different subcomponents of mentalising, and because they use different formats (e.g., vignettes, photographs, and film), each task will also draw upon different non-mentalising skills (Bialecka-Pikul et al., 2021; Warnell & Redcay, 2019). The current review aimed to explore the different formats that have been used in these tasks

(e.g., vignettes, photographs, and film) and the different mentalising and non-mentalising skills that each task draws upon. Awareness of the properties of mentalising tasks is particularly important when studying the relationship between mentalising and language to account for shared task demands. Three out of the five most frequently used tasks are text-based (false belief tasks, the Strange Stories task, and the Hinting Task), and even mentalising tasks using visual stimuli often require verbal responses (e.g., the Reading the Mind in the Eyes Test and Frith-Happé Animations Test). Therefore, any correlations found between mentalising and language could be due to the verbal demands of the tasks rather than any conceptual relationship between mentalising and language. For this reason, the current review explored the mentalising tasks that have been used to investigate the relationship between mentalising and language.

As a first step, the tasks that have been used to operationalise mentalising in adults were collated. To this end, a broad inclusion criteria were used to examine papers published from 2002 to 2022, as initial investigation suggested that few papers cover mentalising and language specifically in adults. Second, a comprehensive taxonomy of the mentalising tasks used to investigate the relationship between mentalising and language was developed. This was done to establish which subcomponents of mentalising each task measures and which non-mentalising skills the tasks draw upon. Finally, the reported correlations between the mentalising tasks and language abilities were examined to understand this relationship. This systematic review addressed three main research questions:

RQ.1) What are the different ways in which mentalising has been operationalised in adult samples?

RQ.2) What are the properties of these mentalising tasks?

RQ.3) What are the reported relationships between mentalising and components of language?

3.3 Method

3.3.1 Eligibility criteria

This systematic review aimed to describe the way in which mentalising has been operationalised when examining the relationship between mentalising and language abilities in neurotypical adults. Research over the last 20 years was examined, as our preliminary analysis of research in the area revealed sparse studies of this relationship in the adult population. The following eligibility criteria were used:

Type of study to be included:

1. Experimental and correlational study designs (no reviews or commentaries, etc.) that included at least one mentalising and one language measure. Intervention studies were included; qualitative studies were not included.
2. Studies that provide original findings regarding mentalising and language.
3. Studies published in the English language.
4. Studies published from 2002 to 2022.

Population/participants:

1. Adults aged 18 and over.
2. Studies with clinical samples (e.g. autism; dementia) were included if the design contained a control group of adult participants without known psychiatric or neurological conditions, severe sensory impairments such as deafness or blindness, or language impairment.

Variables/ measures:

1. As a key aim of this review was to collate reported correlations between mentalising and language skills, any measure of mentalising or language could be a predictor, control, or outcome measure in the study design. Language variables included both measures of oral language (e.g., vocabulary knowledge) and written language (e.g. reading comprehension).

3.3.2 Search strategy

Existing research literature on mentalising in adults published in English between 2002 and 2022 was systematically searched for. The searches were conducted between June and July 2022 using the following databases: PsychInfo, PsyArticles (American Psychological Association), and ERIC (Institute of Education Science). There were three categories of search terms:

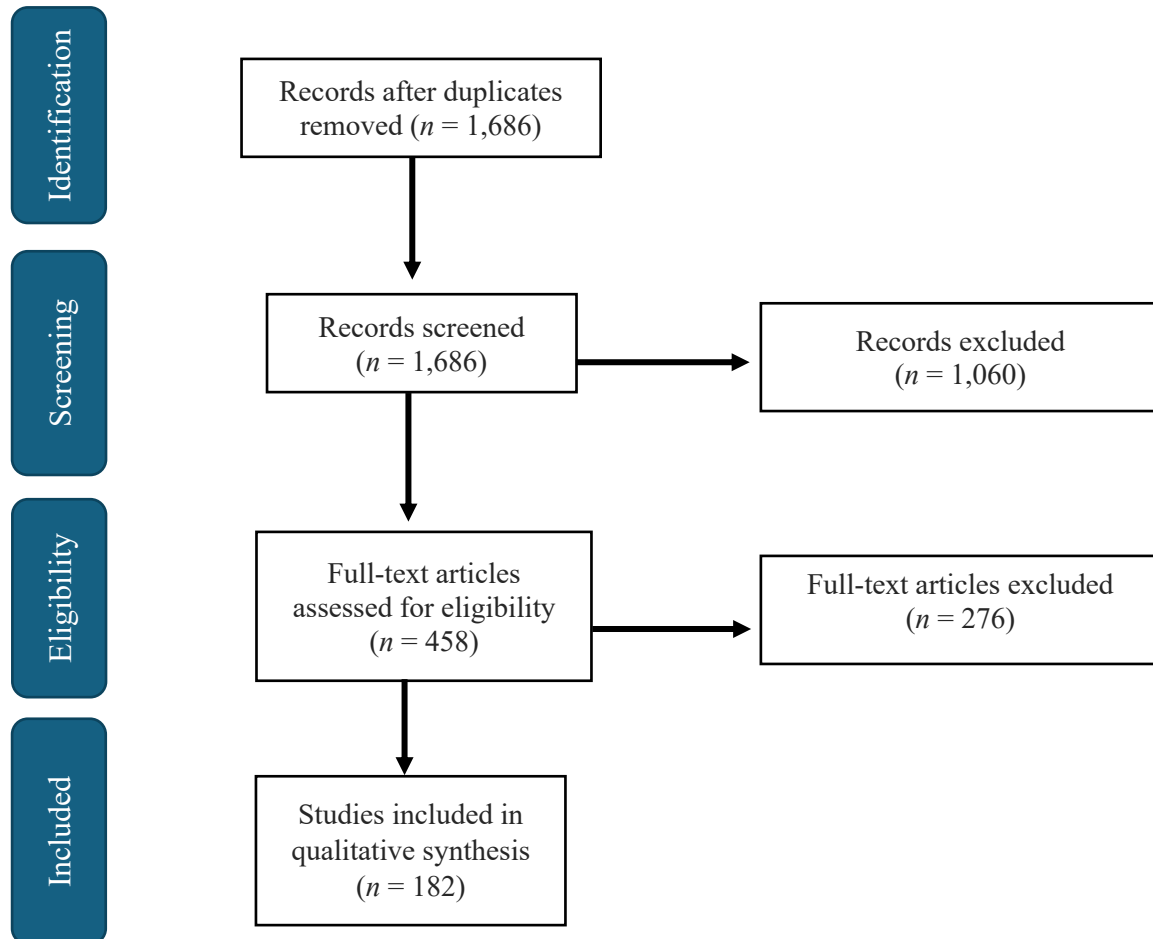
- (1) Mentalising, for which six terms were used (*cognitive empathy* OR *mentalising* OR *mind reading* OR *mindreading* OR *perspective taking* OR *mentalising*);
- (2) Age range, for which two terms were used (*adult** OR *lifespan*);

(3) Language, for which seven terms were used (*comprehension* OR *grammar* OR *language* OR *narrative* OR *pragmatics* OR *reading* OR *vocab**). The categories searched included academic journals, dissertations, and reports, but excluded books. Studies were also hand-searched (i.e., manually checking through databases for eligible papers).

The initial search yielded 1,686 results. After a title-only search, 636 records remained, which were reduced to 458 after the abstracts were screened to ensure they fit the eligibility criteria. Finally, after the full texts were examined, 180 records remained, and two additional papers were found via hand searching, resulting in 182 records in total that included the measurement of mentalising skills in adults (see Figure 3.1). Of these, 49 studies reported correlational statistics between specific mentalising tasks and the subcomponents of language.

Figure 3. 1

Flowchart of the Process of Limiting Publications from the Initial Database Search



3.2.3 Coding

The primary researcher coded each study under the supervision of two researchers who were involved throughout the duration of the systematic review to ensure the level of bias was assessed and minimised. After text and title screening, a randomised selection of 20% of the papers was assigned to two research assistants alongside a SPIDER (sample, phenomenon of interest, design, evaluation, research type) tool, to ensure that all studies were validly selected. The SPIDER tool is a checklist that researchers can follow to examine whether the papers included meet the correct criteria regarding the sample, phenomenon of interest, design, evaluation, and research type. Discrepancies were resolved in meetings between the primary researchers.

The primary coding variables were as follows: (1) participants: studies must have involved neurotypical adults – if a paper focused on a clinical population but included a control group without any known psychiatric or health problems, it was also included; (2) age: participants had to be 18 years old and over – this was checked by calculating the mean age of the sample; (3) measures: papers must have included at least one mentalising measure and one language measure; (4) studies must have provided primary data; and (5) the language variables had to have been categorised (e.g., as vocabulary or oral language comprehension). In the synthesis of the mentalising and language results, correlation coefficients were recorded. Thereafter, the papers were further analysed and categorised to determine which tasks were most frequently used in studies examining mentalising and language in adults over the last 20 years. The following aspects were then explored in more depth: the characteristics of the three most commonly used mentalising measures in the literature (the Reading the Mind in the Eyes Test, the Strange Stories task, and the Frith-Happé Animations Test) (see Section 3.3.2); the operationalisations most frequently used to

investigate the relationship between mentalising and language; and the most common stimuli formats for measuring mentalising (e.g., vignettes, film, and still cartoons). Finally, a taxonomy of the mentalising and non-mentalising properties of the tasks was formed. This systematic review was pre-registered with the [Open Science Framework](#).

3.4 Results

3.4.1 Tasks used to measure mentalising in language

In total, 182 papers were selected after the full papers were reviewed, each containing one or more mentalising tasks. Many of these studies used multiple measures of mentalising: Mentalising was measured 349 times across the 182 studies using 91 different mentalising tasks; 80 of these studies (44%) used only one measure of mentalising. An examination of all the mentalising tasks used across the 182 studies revealed 91 original mentalising tasks. Table 3.2 lists the seven most frequently used tasks, which together account for 68% of all the mentalising tasks used across the 182 studies.

Table 3. 2*The Seven Most Frequently Used Tasks Across the 182 Studies Selected*

Task	<i>n</i>
Reading the Mind in the Eyes Test	95
Strange Stories Task	37
Faux Pas Task	29
The Awareness of Social Inference Test	16
Hinting Task	15
Frith-Happé Animations Test	12
Movie for the Assessment of Social Cognition	10

As Table 3.2 shows, the most frequently used task was Reading the Mind in the Eyes, which represents 52% of the mentalising tasks used over the 182 studies. Amongst the 80 studies that only used one mentalising measure, the most commonly used task was Reading the Mind in the Eyes (26 studies, 33%), and the second and third most common tasks were the Strange Stories task (11 studies, 14%) and the Faux Pas Task (5 studies, 6%), respectively.

3.4.2 Differences in the way mentalising tasks are described across studies

Analysis of the descriptions of the tasks in the selected studies revealed wide variation across studies, most notably the Reading the Mind in the Eyes Test. This section addresses the ways in which three mentalising measures used with adults are described in

the research literature: the Reading the Mind in the Eyes Test, the Strange Stories task, and the Frith-Happé Animations Test). These three tasks were selected for detailed review because, first, the Reading the Mind in the Eyes and Strange Stories tasks were the most commonly used tests to explore the relationship between mentalising and language in adults. Second, the most frequently used visual animated task (Frith-Happé Animations) was commonly used in combination with text-based mentalising tasks.

Although definitional inconsistencies can be found across the literature in relation to several tasks, the definition of the Reading the Mind in the Eyes Test was examined in detail here, because it was the most conspicuous example of conceptual inconsistency in mentalising research.

Reading the Mind in the Eyes Test (mental state attributions from non-verbal cues)

The most frequently used task across the studies was the Reading the Mind in the Eyes Test. This measure is defined in number of ways across studies, including: measuring mentalising (Walsh-Messinger et al., 2019), attribution of complex emotional and cognitive states (Meinhardt-Injac et al., 2020), affective ToM (theory of mind) (Booules-Katri et al., 2019), advanced affective ToM (Demers & Koven, 2015), emotional recognition (Ahmed & Miller., 2011), emotional mentalising (Compère et al., 2016), social-perceptual component of mentalising (Lecciso et al., 2016), perception-based identification task (Herbert et al., 2014), explicit mentalising (Grainger et al., 2018), ToM decoding (Wolkenstein et al., 2011), mental state decoding (Rokita et al., 2021), mental state inference (Turkstra et al., 2008), visual ToM (Spencer et al., 2020), advanced meta-representational skills (Zegarra-Valdivia et al., 2019) and social cognition (Martino et al., 2011).

The wide variety of terms used to describe the Reading the Mind in the Eyes Test is confusing, but this does not necessarily mean there is disagreement over what the task measures. Rather, a lack of agreement regarding how the tasks should be labelled may cause confusion. According to Rosenbaum et al. (2015), the ability to infer one's own mental states and the mental states of others is referred to as theory of mind, mental state attribution, or mentalising, indicating that these terms are used interchangeably. Thus, rather than disagreement about what the task measures, different researchers may use different words to describe the same construct. In the above-mentioned papers, authors do not clearly explain why they labelled the Reading the Mind in the Eyes Test as they did. This may indicate that the authors believed the label they used was not controversial and that the research community agreed on the concepts the task measures. However, other descriptions of the task may indicate a fundamental problem of conceptual disagreement over what the Reading the Mind in the Eyes Test measures.

Specifically, in research, uncertainty currently exists regarding whether the Reading the Mind in the Eyes Test measures both cognitive and affective mentalising or solely affective mentalising. Cognitive mentalising involves making inferences about people's knowledge, intentions, and beliefs, whereas affective mentalising involves making inferences about people's emotions (O'Brien et al., 2011). In contrast, emotion perception is the ability to *recognise* emotions in another human. Research in emotion perception often examines the ability to recognise emotions in other people's faces (Engelmann & Pogosyan, 2013). In order to make a mental state inference about someone's emotional state, we need to first recognise their emotions. Therefore, affective mentalising relies on emotion perception. If the Reading the Mind in the Eyes Test is a test of emotion perception we would expect strong correlations with other emotion perception tasks.

Studies examining the psychometric properties of the Reading the Mind in the Eyes Test have found that the task is moderately associated with measures of emotion perception (Miguel et al., 2017; Olderbak et al., 2015; Seo et al., 2020). In these studies, performance on the Reading the Mind in the Eyes Test was compared with performance on tasks that showed pictures or videos of faces expressing emotion, which participants had to identify using emotional words. However, researchers found that scores on the Reading the Mind in the Eyes Test correlated less strongly with emotional perception tasks than expected ($r = .43^{**}$), suggesting that the Reading the Mind in the Eyes Test is not purely an emotional recognition task (Miguel et al., 2017). The task items contain words that draw on cognitive intentions (e.g., joking and interested) as well as emotional words (e.g., jealous); therefore, the task may cover both cognitive and affective mentalising, which may explain why stronger correlations with emotional processing were not observed in Miguel et al.'s (2017) study. Alternatively, the reason stronger correlations were not observed may be because emotion perception tasks in this study involved stimuli that showed the whole face, not just the eyes (Miguel et al., 2017; Olderbak et al., 2015; Seo et al., 2020). Emotion cues are contained in moving facial muscles, which cause changes to the skin around the eyes, eyebrows, nose and mouth (Neth & Martinez, 2009). Seeing pictures of people's whole faces rather than just the eyes may give an advantage in emotion recognition. In conclusion it remains unclear whether the Reading the Mind Test is purely a measure of emotion perception or not.

One of the main barriers to understanding the properties of the Reading the Mind in the Eyes Test is that the task has low internal consistency (Olderbak et al., 2015; Harkenss et al., 2010; Khorashed et al., 2015; Vellante et al., 2012). This may be because although it was made to measure a single construct, mentalising, it in fact measures a variety of different subcomponents of mentalising (e.g., cognitive and affective) and related skills (visual

processing, working memory) (Eddy & Hansen, 2020; Khorashad et al., 2015; Konrath et al., 2014; Olderbak et al., 2015). Furthermore, items draw upon affective states of difference valences (positive, negative, neutral) (Harkness et al., 2012; Konrath et al., 2014). This may explain why the Reading the Mind in the Eyes Test does not strongly correlate with other tasks measuring single constructs, as it measures a range of different skills (Megías-Robles et al., 2020).

Strange Stories (a task that draws on pragmatic language comprehension)

The second most frequently used task, the Strange Stories task, is operationalised here as a task that draws on pragmatic language comprehension. Among the studies reviewed, there was relatively little variance in the way the task was labelled. It was termed a theory of mind task (Schuwerk et al., 2015), a mentalising task (Nentjes et al., 2015), and a cognitive theory of mind task (Stonnington et al., 2013). However, detailed descriptions of the Strange Stories task revealed more variation across the studies. The task has been described as testing participants' understanding of characters' mental states (Langdon et al., 2014), understanding of why the characters say something they do not literally mean and the intentions underlying the statements (Shaw et al., 2004), understanding of the mental state justifications for story characters' non-literal statements (Kristen et al., 2014), understanding of naturalistic situations and why people behave as they do (Lough et al., 2006), and understanding of the intentions behind the characters' actions (Ahmed et al., 2013). Opinion therefore seems to differ regarding whether the task examines participants' ability to infer characters' mental states, to use mental state information to explain characters' actions, or to understand characters' pragmatic language. In any of these conceptualisations, it is clear that verbal ability is central to successful completion of this task, in addition to – or in interaction with – mentalising skills.

Frith-Happé Animations (attributing intentions to moving shapes)

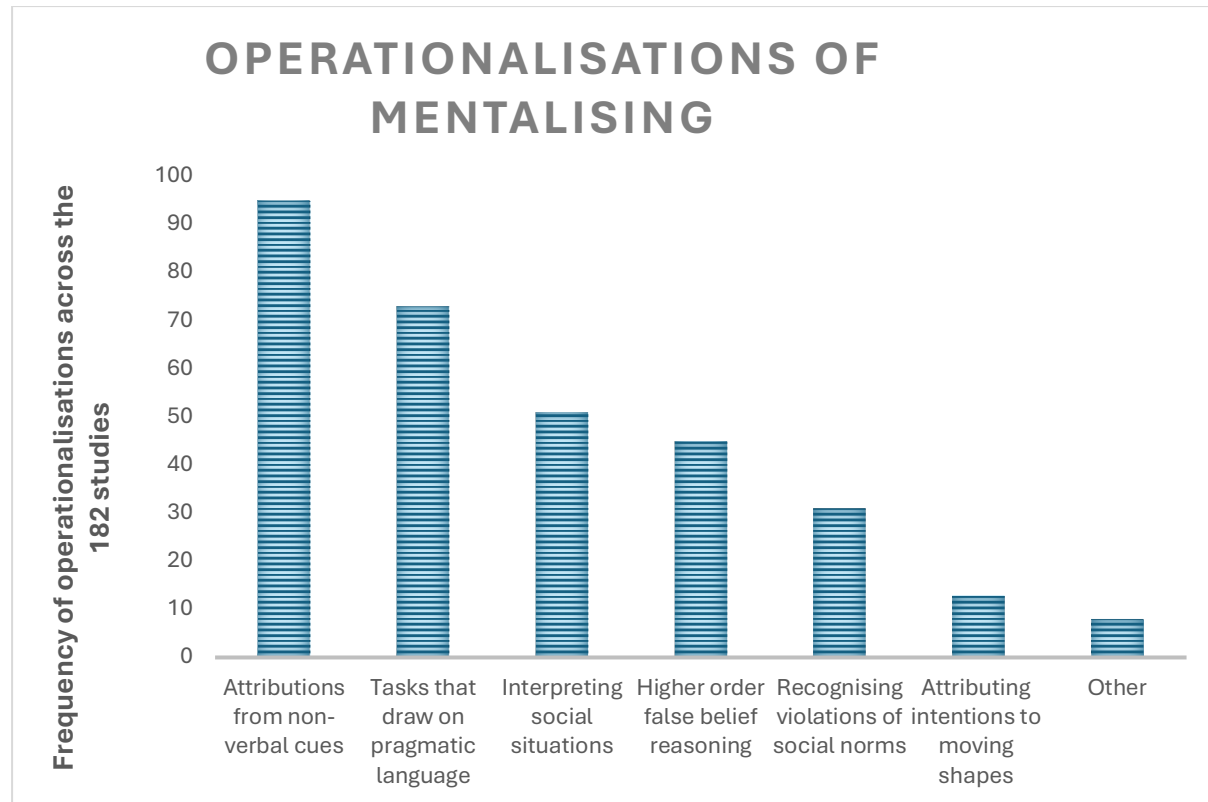
The most frequently used visual animated task, the Frith-Happé Animations Test is operationalised here as attribution of intentions to moving shapes. In the review, small variations were found in how this task was labelled across studies. The task was described as a theory of mind task (Bliksted et al., 2017), a mental state attribution task (Ladegaard et al., 2014), a higher-order mental state attribution task (Stonnington et al., 2013), and a spontaneous mental state attribution task (Nicholson et al., 2019). Although several scores are derived from this test, the ‘intentionality’ score is most clearly related to mentalising. This score was described as measuring the degree of mental state attribution (Bliksted et al., 2017; Stonnington et al., 2013); the degree to which purposeful movements and mental states are described (Horan et al., 2009); the degree to which the participant describes complex, intentional mental states (Lugnegård et al., 2013); and the participant’s ability to interpret the movements of the shapes spontaneously using mental state terms (both cognitive and affective mental states) (Nicholson et al., 2019). The Frith-Happé Animations Test was used much less frequently than the Reading the Mind in the Eyes and the Strange Stories tasks, but subtle variations existed in the descriptions of the intentionality score. The main difference concerned whether participants’ success was due to (a) their ability to attribute mental states to the moving shapes or (b) their ability to use mental state terms to describe these movements effectively.

3.4.3 Operationalisations of mentalising

This review examined the way in which studies operationalised mentalising by clustering different tasks based on the mentalising concepts that the tasks targeted (e.g., false beliefs; pragmatic language). These concepts were further characterised in terms of the format in which tasks were presented; for example, false beliefs were measured using still cartoons, films, vignettes, and picture sequencing. Figure 3.2 illustrates the most frequent ways in which mentalising was operationalised across the 182 studies. The six most commonly used operationalisations were: attributions from non-verbal cues; tasks that draw on pragmatic language, interpreting social situations, higher-order false belief reasoning, recognising violations of social norms, and attributing intentions to moving shapes.

Figure 3. 2

Operationalisations of Mentalising Across the 182 Studies



Note. This figure shows the way mentalising has been operationalised across the 182 studies examining mentalising and language in adults. ‘Other’ refers to self-evaluated mentalising (5), visual perspective taking involving humans (2), and use of mental state terms in expressive language (1).

Across the 182 studies, mentalising in adults was measured 312 times, using 91 different tasks. Many studies used multiple mentalising tasks in their design. Mentalising was most frequently operationalised through attributions from non-verbal cues, accounting for 95 of the (312) total instances (30%). This category contained only one task, namely Reading the Mind in the Eyes.

The second most frequent operationalisation was tasks that draw on pragmatic language (23% of the total measurement instances). This category comprised vignette-based tasks (e.g., the Strange Stories task and the Hinting Task) and films (the Awareness of Social Inference Test; McDonald et al., 2003). The Strange Stories task accounted for 37 out of the 72 tasks categorised as drawing on pragmatic language.

The third most frequent operationalisation was interpreting social situations (16% of the total measurement instances), which involved making inferences about a character's thoughts and feelings (excluding false beliefs) from their actions in the stories. This category included vignette-based tasks (e.g., the Short Story Task), films (e.g., the Movie for the Assessment of Social Cognition), and still cartoons (e.g., Thematic Apperception Test Narrative Task; Murray, 1943).

The fourth most frequent operationalisation was higher-order false beliefs (14% of the total measurement instances), which involved interpreting a character's actions rather than their spoken language. This category consisted of vignette-based tasks (e.g., Sally Anne stories), still cartoons (e.g., The Visual Cartoon Task; Corcoran et al., 1997), films (e.g., UCSF cognitive mentalising test; Shany-Ur et al., 2012), and picture card sequencing.

The fifth most frequent operationalisation was recognising violations of social norms (9% of the total tasks). The Faux Pas Task accounted for 29 out of the 31 tasks in this category. The Edinburgh Social Cognition Test (ESCoT; Baksh et al., 2018), a still cartoon task, was also placed in this category.

The sixth most frequently operationalised category was attributing intentions to moving shapes (4% of the total tasks). The Frith-Happé Animations Test accounted for 10 out of the 13 tasks in this category.

3.4.4 Task formats

There were five commonly used task formats for measuring mentalising; all other formats were used less often. The most commonly occurring format involved vignettes, which 41 (45%) of the total 91 different formats used across the 182 studies. The most frequently used vignette-based mentalising tasks (across 312 total measurement instances) were the Strange Stories task (37), the Faux Pas Task (29), and higher-order false belief stories (22). These three tasks combined represent 28% of the total instances of measuring mentalising across the 182 studies.

Figure 3. 3

The Distribution of Different Tasks Used to Measure Mentalising in the Studies Reviewed

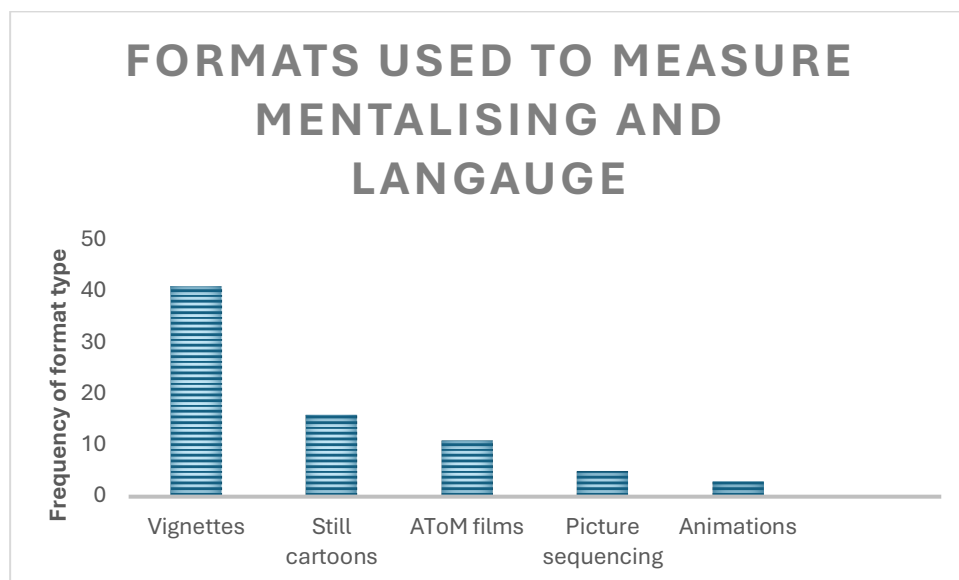
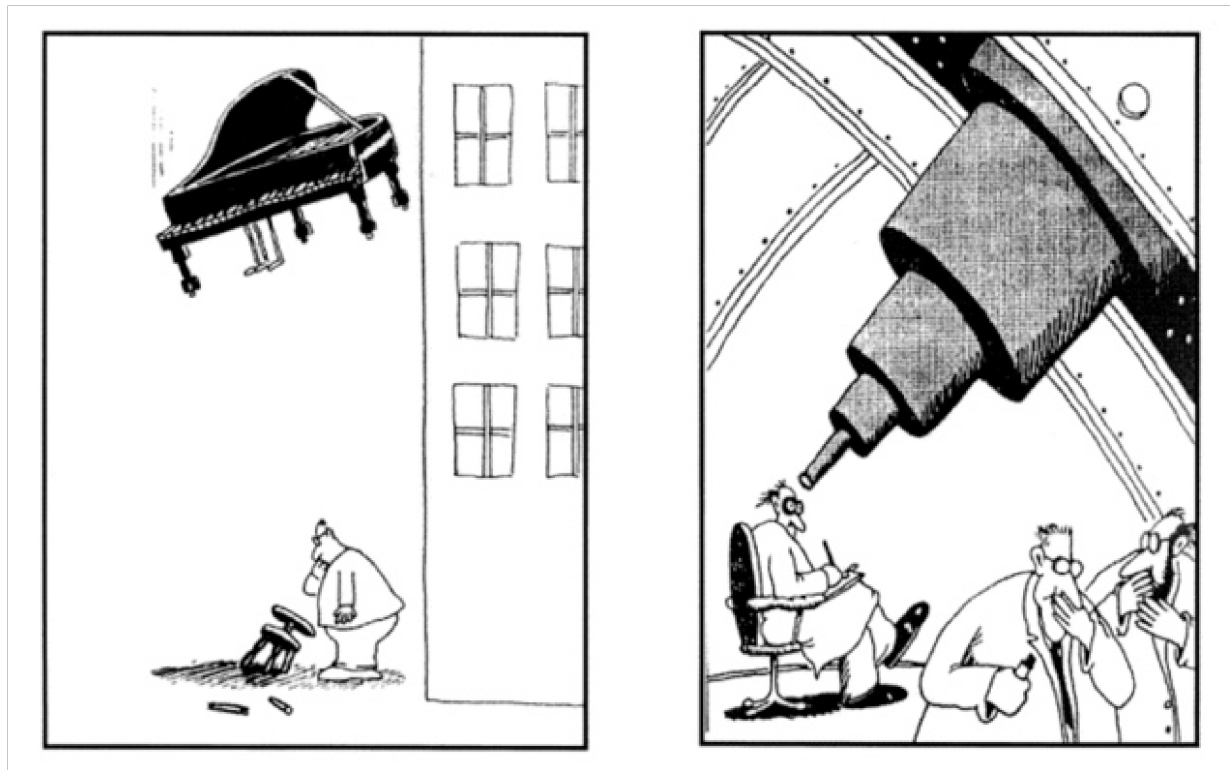


Figure 3. 4

False Belief (Mentalising) Cartoons



Note. Gallagher et al.'s (2000) mentalising still cartoon task requires the participant to attribute ignorance or a false belief to the characters in the cartoon.

The second most frequently used format was still cartoons (see Figure 3.4 for an example). Of the total 91 original mentalising tasks used across the 182 studies, 16 (18%) used a still cartoons-based format. The most frequently used tasks in the still-based-formats category were the Visual Cartoon Test (Corcoran et al., 1997) (3), Brunet et al.'s (2000) AI comic strip (3), and the ESCoT (Edinburgh Social Cognition Test) (2) (Baksh et al., 2018).

The third most frequently used mentalising format was film. Of the total 91 mentalising tasks used across the 182 studies, 11 (12%) used a film-based format. The most frequently used film-based task was the Movie for the Assessment of Social Cognition (8)

(Movie for the Assessment of Social Cognition; Dziobek et al., 2006), followed by the Reading the Mind in Films Task (3) (RMFT; Golan et al., 2006).

3.4.5 Developing a taxonomy of mentalising tasks for adults

A key aim of the review was to form a taxonomy of the mentalising tasks used to measure mentalising in adults. As a large number of tasks (91) were found across the studies included in this review, reviewing the mentalising and non-mentalising properties of all the tasks would be impractical. Instead, a selection was chosen for detailed examination, with a focus on the most commonly used tasks. As these tasks are used most regularly to measure mentalising, researchers must understand the properties of these tasks and the differences between them before designing research projects. Although these tasks all differed in some respects, some were conceptually and practically similar. For instance, 12 different tasks used still cartoons; five of these still cartoons, such as Griffin's (1999) Humorous Cartoon Task and Langdon's (2010) Joke Appreciation Task, used funny still cartoons designed for adults (e.g., Far Side comics). These tasks required participants to explain the humour, and to do this successfully, they had to understand the cartoon characters' false beliefs (Langdon et al., 2014).

To conduct a detailed examination of a selection of the tasks, the number was reduced from 91 to 20 (see Table 3.3). This was done by grouping the tasks, first, in their operational categories (e.g., false belief reasoning and interpreting pragmatic language) and second, within each operational category according to the format used (e.g., still cartoons or vignettes). If there were multiple occurrences of tasks using the same format within each operational category, only the most frequently used task was extracted for the taxonomy. In addition, some tasks were selected that may be suitable alternatives to the most frequently

used verbal tasks (e.g., Strange Stories), which could help reduce shared task demands in mentalising and language research. These were the film-based tasks, namely the Silent Films Task (Devine & Hughes, 2019) and the Adult-Mentalising Task A-ToM (Brewer et al., 2017), which may have been used less frequently, as they have been designed more recently. The Hinting Task was also included as the second most frequently used pragmatic task after Strange Stories because it was the fifth most frequently used task across all studies. This meant that in each operational category, there were between one and four mentalising tasks. Table 3.3 presents a taxonomy of the 20 tasks.

Table 3. 3*Tasks Examined in the Taxonomy*

Tasks	Format of Task	Operationalisations Within 91 Original Tasks		Count/Percentage of Task Use Across the 182 Studies	
		<i>n</i>	%	<i>n</i>	%
Higher-order false beliefs		39	43.86		
<i>The Visual Cartoon Test</i>	<i>Still cartoons</i>			2	0.63
<i>Rowe's (2001) False Belief Stories</i>	<i>Vignettes</i>			2	0.63
<i>Sally Anne Films</i>	<i>Film</i>			3	0.94
<i>Brune's (2003) Picture Card Sequencing</i>	<i>Picture card sequencing</i>			3	0.94
Interpreting Social Situations		21	23.08		
<i>The Short Story Task</i>	<i>Vignettes</i>			1	0.32
<i>Movie for the Assessment of Social Cognition</i>	<i>Film</i>			10	3.16
<i>Spontaneous Mentalising Protocol</i>	<i>Film</i>			1	0.32
<i>Reading the Mind in the Voice</i>	<i>Audio</i>			2	0.63

Tasks	Format of the Task	Operationalisations Within 91 Original Tasks		Count/ Percentage of Task Use Across the 182 Studies	
		<i>n</i>	%	<i>n</i>	%
Tasks That Draw on Pragmatic Language		12	13.19		
<i>Strange Stories</i>	<i>Vignettes</i>			37	11.71
<i>The Hinting Task</i>	<i>Vignettes</i>			15	4.75
<i>Adult-Mentalising Task</i>	<i>Film</i>			1	0.32
<i>Versailles-Situational Intentional Reading</i>	<i>Film</i>			2	0.63
Attributions From Non-Verbal Cues		1	1.10		
<i>Reading the Mind in the Eyes</i>	<i>Photographs</i>			95	30.06
Visual Perspective Taking (Involving Humans)		3	3.30		
<i>The Director's Task</i>	<i>Computerised game</i>			2	0.63
Attributing Intentions to Moving Shapes		4	4.40		
<i>The Frith-Happé Animations</i>	<i>Animations</i>			11	3.48

Tasks	Format of the Task	Operationalisations Within 91 Original Tasks		Count/ Percentage of Task Use Across the 182 Studies	
		<i>n</i>	%	<i>n</i>	%
<i>Edinburgh Social Cognition Test (ESCoT)</i>	<i>Animations</i>			2	0.63
Use of Mental State Terms in Expressive Language		1	1.10		
<i>ToM Story</i>	<i>Narrative production</i>			1	0.32
Self-Evaluated Mentalising		8	8.79		
<i>Spontaneous Mental State Language</i>	<i>Interview</i>			1	0.32
<i>THOM.a.s</i>	<i>Questionnaire</i>			1	0.32

Note. This table displays the most frequently used tasks in each operational category within the 91 original tasks and the number of times each individual task was used across the 182 studies.

Table 3. 4 Taxonomy of mentalising task properties.

	Higher-Order False Belief Reasoning				Interpreting Social Situations				Tasks That Draw on Pragmatic Language	
	Cartoon	Vignette	Film	Picture Sequencing	Vignette Task	Film Task	Film Task	Audio Task	Vignette Task	Vignette Task
	The Visual Cartoon Task	Rowe's False Belief	UCSF Cognitive ToM Test	Brune Cartoon Pictures	The Short Story Task	MASC	Silent Films Task	Reading the Mind in the Voice	Strange Stories	Hinting Task
Cognitive mentalising	X	X	X	X	X	X	X		X	X
Affective mentalising					X	X	X	X	X	
Emotional processing					X		X	X		
Tracking of visual perspectives										
Recognising violations of social norms						X				
Use of mental state language					X		X			
Emotional attributions					X		X			
Attributing mental states to humans					X		X	X	X	
False beliefs	X	X	X	X		X	X		X	X
Deception/lies	X			X	X	X			X	
Jokes	X				X				X	
Pragmatics (hints)										X
Pragmatics (indirect requests)					X					X
Pragmatics (figure of speech)					X				X	
Pragmatics (irony)					X				X	
Pragmatics (other)					X	X			X	

Table 3.4 cont. Taxonomy of mentalising task properties (non-mentalising).

	Higher-Order False Belief Reasoning				Interpreting Social Situations				Tasks That Draw on Pragmatic Language	
	Cartoon	Vignette	Film	Picture Sequencing	Vignette Task	Film Task	Film Task	Audio Task	Vignette Task	Vignette Task
	The Visual Cartoon Task	Rowe's (2001) False Belief	UCSF Cognitive Mentalising Test	Brune (2003) Cartoon Pictures	The Short Story Task	MASC	Silent Films Task	Reading the Mind in the Voice	Strange Stories	Hinting Task
Face processing (human)										
Visual scene processing (cartoon)	X			X						
Visual scene processing (film)			X			X	X			
Vocabulary			X					X		
Reading comprehension					X				X	X
Listening comprehension		X	X			X		X		
Visual processing				X						
Narrative production (make up a story)				X						
Expressive language		X	X		X	X	X		X	X
Physical causality	X	X	X							

Table 3.4 cont. Taxonomy of mentalising task properties (mentalising).

	Tasks That Draw on Pragmatic Language		Attributions From Non-Verbal Cues	Visual Perspective Taking (Humans)	Attributing Intentions to Moving Shapes	Recognising Violations of Social Norms		Mentalising in Expressive Language	Self-Evaluated Mentalising (Interview Questions)	
	Cartoon	Vignette	Film	Picture Sequencing	Vignette Task	Vignette Task	Film Task	Audio Task	Interview Task	Vignette Task
	V-SIR	Adult-Mentalising Task	Reading the Mind in the Eyes Task	Director's Task	Frith-Happé Animations	Faux Pas Task	Edinburgh Social Cognition	ToM Story	Spontaneous Mentalistic Language	THOM.a.s (Open-Ended Questions)
Cognitive mentalising	X	X		X	X	X	X	X	X	X
Affective mentalising	X		X				X	X	X	X
Emotional processing	X		X						X	X
Self/other attributions									X	X
Tracking of visual perspectives				X						
Recognising violations of social norms						X	X	X		
Use of mental state language					X			X	X	X
Emotional attribution			X							
Attributing mental states to humans	X	X								X
Attributing mental states to objects					X					
False beliefs		X							X	
Deception/lies	X	X								
Jokes	X	X								
Pragmatics (hints)	X									
Pragmatics (indirect request)	X									
Pragmatics (irony)		X								

Table 3.4 cont. Taxonomy of mentalising task properties (non-mentalising).

	Pragmatic Language		Attributions From Non-Verbal Cues	Visual Perspective Taking (Involving Humans)	Attributing Intentions to Moving Shapes	Recognising Violations of Social Norms		Use of Mentalising in Expressive Language	Self-Evaluated Mentalising (Interview Questions)	
	Cartoon	Vignette	Film	Picture Sequencing	Vignette Task	Film Task	Film Task	Audio Task	Interview Task	Vignette Task
	V-SIR	Adult-Mentalising Task	Reading the Mind in the Eyes Task	Director's Task	Frith-Happé Animations	Faux Pas Task	Edinburgh Social Cognition Test	ToM Story	Spontaneous Mentalistic Language	THOM.a.s (Open-Ended Questions)
Face processing (human)										
Visual social scene processing (cartoon)							X			
Visual social scene processing (film)	X	X								
Vocabulary			X				X			
Reading comprehension						X				X
Listening comprehension	X	X		X			X		X	
Visuospatial reasoning				X	X					
Narrative production (make up a story)								X	X	
Expressive language	X	X			X	X	X	X	X	X
Physical causality					X				X	

3.4.6 Associations between mentalising and language

A secondary aim of this review was to investigate the relationship between measures of mentalising and language in adults. As research in this area is limited, a search strategy was employed that included a broad range of language measures. Hence, the papers selected for inclusion in this review were not all primarily designed to study the relationship between mentalising and language; in these papers, language measures were often used as control measures or to screen participants to ensure they were eligible for inclusion in studies. This may partially explain why only a few correlation coefficients were reported: 41 studies (22.5%) reported correlation coefficients between mentalising and language across the 182 studies included in this review. Correlations between mentalising and language were assessed for the five most commonly used mentalising tasks (the Reading the Mind in the Eyes Test, the Strange Stories task, the Faux Pas Task, the Awareness of Social Inference Test, and the Hinting Task), as most of the reported correlations pertained to these measures. However, as no correlations were reported between either the Hinting Task or the Awareness of Social Inference Test and language abilities, these tasks were removed from the analysis, and the sixth most frequently used task (the Frith-Happé Animations Test) was included instead. A total of 43 correlation coefficients were reported between these four mentalising measures and language abilities, 23 of which were reported for the Reading the Mind in the Eyes Test, 14 for the Strange Stories task, five for the Faux Pas Task, and two for the Frith-Happé Animations Test.

Although there were a limited number of correlation coefficients reported, the results indicate positive correlations between the most commonly used mentalising measures and a range of language abilities (discussed below). These subcomponents of language include verbal fluency, word reading, vocabulary, inductive reasoning, the ability to understand and

use spoken language, pragmatic language abilities, paragraph comprehension, and reading comprehension. These correlations were reported in relation to mentalising tasks that used different formats (e.g., vignettes, photographs, and animations). Tables 3.5 to 3.8 describe the reported correlations between four of the most commonly used mentalising measures and a range of language abilities. The sample size (or size of the typically developing control group) in each study is listed; many of the studies only used a small sample size, probably because the typically developing adult participants were often included as control groups in clinical studies. The small sample size would partially explain the variation in performance in individual mentalising tasks and specific language abilities (e.g., verbal fluency) from one study to another, as correlation estimates in small samples are often less accurate than larger samples (Schönbrodt & Perugini, 2013).

Reading the Mind in the Eyes Test (attributions from non-verbal cues)

Table 3. 5 - Reported Correlations Between the Reading the Mind in the Eyes Test and Language Abilities

Journal Reference	Verbal Fluency	Word Reading	Vocab.	Pragmatic Ability		Reading Comp.	Mean Age	Sample Size	First language of sample
Muller et al. (2010)	.23 (.05)						37	15	French
Rigby et al. (2016)	.51 (.26)	.51** (.26)					31	41	English
Ahmed & Miller (2011)	.23 (.05)						19	123	English
Peterson & Miller (2012)			.49** (.24)	.43* (.18)	figurative language		19	42	English
Lugnegård et al. (2013)			.20 (.04)				28	50	Swedish
Whitton et al. (2013)	.52** (.27)						40	44	English
Kristen et al. (2014)							29	20	German
Hezel et al. (2014)		.23* (.05)	.32 (.10)				20	40	English
Demers et al. (2015)			.22* (.05)				19	86	English
Olderbak et al. (2015)			.54*				33	483	German
Bambini et al. (2021)				.45** (.20) .41** (.17)	General pragmatics implicatures		82	54	Italian
Camodeca (2019)			.23 (.05)				33	147	English
Meinhardt-Injac et al. (2020)			.52** (.27)				20	293	German
Adams et al. (2020)	.43 (.18)	.43 (.18)					64	53	English
Baksh et al. (2021)		.30 (.09)					37	38	English
Jackson (2022)						.23** (.05)	40	220	English
Fastame & Carta (2021)	.32 (.10)		.52 (.27)				81	56	Italian
Giordano et al. (2019)	.30 (.09)	.23** (.05)					22	9	Italian
Weighted means	.48**	.39**	.48**	.43**		.23**			

* $p < .05$, ** $p < .001$, the effect sizes are placed in parenthesis below each correlation.

The results suggest that the Reading the Mind in the Eyes Test correlates with several language abilities (see Table 3.5), indicating that the mentalising properties (attributions from non-verbal cues), or the non-mentalising properties of the task or both are associated with language abilities. The Reading the Mind in the Eyes Test likely draws on non-mentalising abilities (e.g., vocabulary knowledge), as the participant must have an understanding of the mental state words provided to describe the eyes, to accurately identify which mental state the pictures of eyes are portraying in each item. In the 2001 revision of the Reading the Mind in the Eyes Test (which is the most commonly used version in contemporary research), Baron-Cohen and colleagues attempted to make it harder to identify the mental states in the task. To do this, more complex words were added in order to increase the difficulty of the task. In addition, the emotional valence (the degree to which an emotion is positive or negative) of the three distractor words was matched to the emotional valence of the target word. For example, on one item, the target word is “serious”, and the distractors were “ashamed”, “alarmed”, and “bewildered”. These amendments may have made it harder to identify the correct mental state, as the task may now draw more heavily on the participants’ knowledge of mental state words. This may explain why small to moderate correlations were found between the Reading the Mind in the Eyes Test and vocabulary ($r = .20-.54$, weighted mean = $.48^{**}$) and Word Reading ($r = .23-.54$, weighted mean = $.39^{**}$). Word reading (or word decoding) measures test individuals’ understanding of orthography (correct spelling of words) and phonology (knowledge of letter-speech sounds). Research demonstrates that when individuals learn to decode words, they increase their vocabulary knowledge by remembering the meaning of words as they are (Nation & Snowling, 1998; Perfetti et al., 2010). Furthermore, Peterson and Miller (2012) also included a facial recognition task that did not explain any variance on the task, and the authors suggest that the variance unaccounted for by vocabulary can be accounted for by social-

perceptual processes, which would indicate that the Reading the Mind in the Eyes Test is a relatively reliable test of mentalising, if vocabulary knowledge is controlled. However, overall these findings suggest that people with better vocabulary may perform better on the Reading the Mind in the Eyes Test.

On the other hand, it may be that the Reading the Mind in the Eyes Test correlates with language due to the mentalising properties of the task. At a young age (eight years old) children's use of mental state words has been moderately correlated with measures of cognitive and emotional mentalising (Grazzani & Ornaghi, 2012). It may be that children who have a better understanding of mental state words can engage in more complex social interactions and conversations, which helps them link thoughts to behaviours in a more sophisticated way, and, in turn improves their mentalising abilities (Nelson, 2005). In adulthood, reading literature may improve mentalising abilities by letting readers simulate interactions between characters as they read (Mar & Oatley, 2008). A recent study has reported a significant correlation between reading comprehension and the Reading the Mind in the Eyes Test; however, again, vocabulary elements of the task could explain this association (Jackson, 2022). As mentalising has been shown to predict reading comprehension concurrently and longitudinally in children, Jackson (2022) explored whether a mutually reinforcing relationship existed between reading comprehension and mentalising in adults). In this study, a small significant correlation ($r = .23^{**}$) was observed between reading comprehension and mentalising, using the Reading the Mind in the Eyes Test, which the authors argue shows initial evidence of a mutually reinforcing relationship between reading comprehension and mentalising. However, like Bambini et al. (2021), the author used no verbal control measure, and it is thus unclear how much of the variance can be explained by the task's mentalising properties.

The Reading the Mind in the Eyes Test has also been associated with pragmatic language (weighted mean = .48**), which could be due to the tasks' language demands (mental state vocabulary) and/or the mentalising properties of the task. Bambini et al. (2021) reported a significant correlation between the Reading the Mind in the Eyes Test and general pragmatic abilities ($r = .45^{**}$) and implicatures ($r = .41^{**}$) in the Italian language. The general pragmatic abilities test is a tool to assess expressive and receptive pragmatic skills that were developed and validated for the Italian-speaking population (Arcara & Bambini, 2016). It examines figurative language, mental state humour and mental state reasoning in narratives. In Peterson and Miller's (2012) study in the English language, they also found a significant correlation between mentalising and figurative language (.43, $p < .05$), so it may be that the relationship between mentalising and figurative language is consistent across the English and Italian languages.

Bambini et al. (2021) ran a multiple regression that suggested higher scores on the Reading the Mind in the Eyes Test predicted higher scores on both the general pragmatic abilities task and the implicatures task, in a model that accounted for age, gender, education level and executive functions. Indicating that mentalising may support these pragmatic language abilities. However, the regression did not control for vocabulary, which may have explained some variance between the Reading the Mind in the Eyes Test and the pragmatic language tasks (Meinhardt-Injac et al., 2020). Additionally, both pragmatic language tasks were presented via a mixture of oral and written text. It is possible that the ability to read people's thoughts and emotions via the eyes is more important when interpreting pragmatic language when it is presented orally than when presented via written text. Yet Peterson and Miller (2012) also found significant correlations between the Reading the Mind in the Eyes Test and figurative language when the figurative language task was presented via written text only, suggesting that mentalising may support the comprehension of written pragmatic

language also. Overall, these results suggest that there may be a relationship between mentalising, as measured by the Reading the Mind in the Eyes Task, and pragmatic language ability.

Strange Stories (task that draws on pragmatic language)

Table 3. 6

Reported Correlations Between the Strange Stories Task and Language Abilities

Journal Reference	Verbal Fluency	Word Reading	Vocab	Pragmatic Ability		Paragraph Comp.	Mean Age	Sample Size	First language of sample
Gavilán & García-Albea (2011)			.17 (.03)	.13 (.02) .16 (.03) -.01 (.00)	Metaphors Irony Proverbs	.31 (.10)	42	22	Spanish
Gibbons et al. (2007)		.22 (.05)					58	16	English
Ahmed & Miller (2011)	.22* (.05)						19	123	English
Ahmed & Miller (2013)				.40* (.16)	Proverbs		77	43	English
Kristen et al. (2014)			.60* (.36)				20	40	German
Lecce et al. (2019)	.16 (.03)		.29* (.08)				68	72	Italian
Cavallini et al. (2013)	.61** (.37)		.30* (.09)				55	86	Italian
Giordano et al. (2019)	.22* (.05)		.28** (.07)				22	9	Italian
Weighted means	.40**	.22	.38**	.24*		.31			

* $p < .05$, ** $p < .001$, the effect sizes are placed in parenthesis below each correlation.

The results suggest that the Strange Stories task correlates with several language abilities (see Table 3.6), indicating that the mentalising properties (tasks that draw on pragmatic language) or the non-mentalising properties (verbal abilities), or both, are associated with language abilities. Correlations were found between the Strange Stories task and language measures, including verbal fluency, vocabulary, paragraph comprehension, and pragmatic language. As the Strange Stories task is a text-based mentalising measure, these correlations may have associations with both the verbal and mentalising properties of the task. In the Strange Stories task, participants read a vignette about a social situation and are then asked why a character said or did something. In the taxonomy, this task was operationalised as a ‘task that draws on pragmatic language’, but it also draws on participants’ ability to infer the mental states and intentions of characters, as inferring others’ intentions is key to understanding pragmatic language (Grice, 1975). Notably, the task did correlate with measures of pragmatic language ability ($r = -.01-.40$, weighted mean $.24$, $p < .001$). As participants read vignettes in the Strange Stories task, the task draws upon reading comprehension, and consequently, language abilities that have been shown to support reading comprehension correlate with the task (Kim, 2017). The results indicate that a direct link exists between the Strange Stories task and reading, as paragraph comprehension was found to correlate with the task ($r = .31$). However, this correlation was not significant. The Strange Stories task also correlated with abilities associated with reading, e.g., vocabulary ($r = .17-.60$, weighted mean $= .38$, $p < .001$), as well as tasks that correlate highly with vocabulary, such as verbal fluency ($r = .16-.61$, weighted mean $= .40$, $p < .001$) (Bell et al., 2001; Perfetti et al., 2010). Together, these results suggest that the Strange Stories task draws on language abilities underpinning reading comprehension (e.g., vocabulary), and pragmatic abilities.

As the Strange Stories task has been described as a task that draws on pragmatic language, one would have expected strong correlations between the task and pragmatic abilities, but the correlations varied in strength (Bosco et al., 2018b; Osterhaus & Bosch, 2022; White et al., 2009). Although, it must be noted that there were only two papers that reported correlations between the Strange Stories and pragmatic language. However, in the papers reviewed, the reasons different associations were reported between the Strange Stories task and proverb interpretation may be due to the differences in the format of the Strange Stories task, the properties of the proverb tasks used, and language differences. Gavilán and García-Albea (2011) reported small positive correlations or no correlation between the Strange Stories task and irony ($r = .16$), metaphors ($r = .13$), and proverbs ($r = .01$), while Ahmed and Miller (2011) found a significant moderate correlation between the Strange Stories task and proverbs ($r = .40, p < .05$). The reason Gavilán and García-Albea (2011) found small correlations may be because they only used three false belief items from the Strange Stories task. Therefore, on the one hand, adults' ability to understand the multiple mentalising concepts drawn upon by the Strange Stories task may be associated with understanding of non-literal speech rather than merely understanding false beliefs. This argument is supported by the findings of Ahmed and Miller (2011), who used the full range of items that test other theory-of-mind concepts (e.g., white lies, double bluffs, and persuasion) and found a moderate correlation between the Strange Stories task and proverbs. On the other hand, there may be some overlap between mentalising concepts, other than false beliefs, and the pragmatic language tasks. For example, the persuasion item requires participants to understand that a threat made by the protagonist (to drown the cats) is not literally true but is made to 'force' a potential buyer of the cats to take them off her hands. Specifically, participants are asked, 'Why did she say that?' Understanding the intentions behind non-literal speech in this item would draw upon pragmatic skills; thus, both this item

and the pragmatic language task may recruit the same skills to some extent. In addition, the proverb tasks used in these two studies had different formats, which might also partially explain these differing results. Gavilán and García Albea's (2011) Spanish language study used a paradigm where the participants heard a story, then the proverb, and then a forced choice selection of interpretations, whereas in Ahmed and Miller's (2011) English language study, participants were given common and uncommon proverbs and asked about their meaning with no context. Alternatively, it may be that mentalising is recruited differently in the two languages (Spanish vs English). These methodological differences may explain why significant moderate correlations were found between the Strange Stories and proverbs in the Ahmed and Miller's (2011) study but not the Gavilán and García Albea's (2011) study.

Faux Pas Task (recognising violations of social norms)

Table 3. 7

Reported Correlations Between the Faux Pas Task and Language Abilities

Journal Reference	Verbal Fluency	Word Reading	Vocabulary	Mean Age	Sample Size	Effect Size	First language of sample
Muller et al. (2010)	.47 (.22)			37	15		French
Ahmed & Miller (2011)	.07 (.01)			19	123		English
Xi et al. (2013)	.17 (.03)		.35 (.12)		20		Mandarin
Meinhardt-Injac et al. (2020)				20	293		German
Radecki et al. (2019)		.59** (.35)		74	90		English
Weighted means	.14	.59**	.35				

* $p < .05$, ** $p < .001$, the effect sizes are placed in parenthesis below each correlation.

Five studies reported correlations between the Faux Pas Task and measures of language. In this task, the participant must read a narrative about a social situation where a violation of social norms has occurred, and they must then identify the faux pas in question. As the Faux Pas Task is a text-based measure of mentalising, it draws on reading comprehension and has been reported to correlate with skills associated with reading comprehension, such as vocabulary ($r = .35$), and other language skills such as and reading ($r = .59$). With regard to verbal fluency scores, Muller et al. (2010) recorded a higher

correlation between the Faux Pas Task and verbal fluency than Ahmed and Miller (2011) and Xi et al. (2013); however, this may be due to the small sample size used ($n = 15$) in Muller's study.

Frith-Happé Animations (recognising violations of social norms)

Table 3. 8

Reported Correlations Between the Frith-Happé Animations Test and Language Abilities

Journal Reference	Verbal Fluency	Vocabulary	Mean Age	Sample Size	Effect Size	First language of sample
Rigby et al. (2016)	.32* (.10)		31	41		English
Lugnegård et al. (2013)		.20* (.04)	28	50		Swedish
Weighted means	.32*	.20*				

* $p < .05$, ** $p < .001$, the effect sizes are placed in parenthesis below each correlation.

Two studies reported correlations between the Frith-Happé Animations Test and language abilities; these were verbal fluency ($r = .32^*$) and vocabulary ($r = .20^*$). In the Frith-Happé Animations Test, the participant watches an interaction between two moving shapes and must attribute human intentions to the interaction (see Chapter 4, Section 4.2, for a complete description of the task). The participant must write a description of the interaction, and the task is scored by recording the amount of intentional language used. The task involves verbal elements and expressive written language, which may explain the association between verbal fluency and vocabulary with the task.

Conclusion

The reported correlations between the five mentalising tasks and language suggest that all five mentalising tasks draw upon aspects of language. This review highlights the fact that both story-based reading tasks (the Strange Stories task and the Faux Pas Task) and tasks using visual stimuli (the Reading the Mind in the Eyes Test and the Frith-Happé Animations Test) draw on language skills. To account for the language properties of the tasks when studying the relationship between mentalising and language, suitable controls should be added (e.g., vocabulary knowledge). Also, more than one mentalising task should be included to compare which aspects of mentalising are related to the phenomenon of interest (see taxonomy, Section 3.4.5). This way, researchers can increase their levels of certainty regarding associations between subcomponents of mentalising (e.g., attributions from non-verbal cues), and related abilities (e.g., language).

3.5 Discussion

This systematic review had three main aims. The first aim was to collate the mentalising tasks that have been used to study the relationship between mentalising and language in adults since 2002 and to review the different ways in which mentalising has been operationalised. Second, it sought to examine the properties of these tasks to form a taxonomy of the individual basic components of the mentalising tasks. The third aim was to investigate reported correlations between mentalising and language.

3.5.1 Measurement and conceptualisations of mentalising

One of the main findings from the systematic review is that mentalising has been measured and conceptualised in many different ways in studies with adult participants. The mentalising concepts measured in these tasks are quite different, which could explain the lack of coherence between different mentalising tasks reported in recent reviews (Białecka-Pikul et al., 2021; Warnell & Redcay, 2019). The five most frequent operationalisations of mentalising, in order, were: mental state attributions from non-verbal cues (e.g., Reading the Mind in the Eyes), tasks that draw on pragmatic language (e.g., Strange Stories), interpreting social situations (e.g., the Short Story Task), higher-order false belief reasoning (e.g., The Theory of Mind Picture Stories Task), and recognising violations of social norms (e.g., the Faux Pas Task). There is a certain amount of overlap between the mentalising concepts drawn upon in each of these groupings. For instance, the Strange Stories task draws on pragmatic language comprehension but may also test a person's ability to interpret social situations. To establish clear links between mentalising and associated skills, such as reading comprehension, researchers must use precise definitions of the mentalising concepts targeted by specific tasks, rather than using diverse mentalising tasks interchangeably.

A second finding from this systematic review is that mentalising tasks have been presented using stimuli in various formats, which could draw upon different non-mentalising skills. The range of non-mentalising skills covered by each task could also contribute to the lack of coherence found between mentalising tasks. Within each operational category (e.g., higher-order false beliefs), a diverse set of formats was used across tasks (e.g., vignettes; still cartoons). The five formats most frequently employed to present mentalising-relevant information to participants were, in order, vignettes, films, still cartoons, animations, and self-report questionnaires. It seems likely that at least some of the association between

mentalising and subcomponents of language is due to verbal demands of the tasks (e.g., vocabulary in the Reading the Mind in the Eyes Test; pragmatic language in Strange Stories). The different formats in which mentalising tasks are presented will likely draw upon other cognitive skills (e.g., executive functions) not addressed in this review. Consequently, it is important to acknowledge that non-mentalising abilities may partially explain associations between mentalising concepts and related abilities (e.g., reading comprehension). For example, recognising violations of social norms can be investigated using a vignette-based format (e.g., the Faux Pas Task) or a still cartoon task (e.g., ESCoT). A correlation between the Faux Pas Task and language, but not between the ESCoT and language, could suggest that the verbal demands of the Faux Pas Task may contribute to this association. In future, research designs employing multiple tasks that address the same mentalising concepts but use different formats may be a useful way to establish relationships between mentalising concepts and associated abilities.

A third finding of this systematic review is that individual differences in mentalising in adults seem to be robustly related to verbal abilities, regardless of (a) the mentalising task used and (b) the language subcomponent (e.g., vocabulary, reading, pragmatics). Even mentalising tasks that do not draw heavily on language (e.g., the Reading the Mind in the Eyes Test and the Frith-Happé Animations Test) show positive associations with language and, most consistently, vocabulary knowledge. Notably, substantially overlapping task demands may exist between some tests of mentalising (e.g., Strange Stories) and tests of pragmatic language. Several factors could explain the variation in effect sizes between studies, including the sample size, language in which tests were administered (English, Spanish, etc.), and variation in the way in which mentalising tasks were delivered (e.g., in full vs short-form versions). Overall, this review suggests that an association exists between verbal skills and mentalising tasks through the lifespan.

There were a few factors that varied in the studies reporting positive correlations between individual differences and verbal abilities. As these studies were run in different countries, the first language of the samples differed. This is relevant, as language and culture have been shown to affect the development of mentalising across the lifespan (Henry et al., 2013). Culture not only shapes how we communicate but also affects how we process emotions (Fung et al., 2008; Stanley et al., 2013). Recent research has shown that, after controlling for age, gender and education, culture explained 25% of variance when recognising violations of social norms (Faux Pas Task) and 20% of recognition of emotions in human faces (Quesque et al., 2022). These findings are supported by Yong et al. (2022), who found that the ability to recognise violations of social norms declined with age in Malaysia, but age effects were not found in Britain. The researchers suggest that this difference in mentalising ability has arisen due to differences in cultural norms; specifically, in Asian cultures, there is less of a tradition of making explicit cultural judgements about other individuals than in the West. Similarly, differences in the type of emotion people are drawn to differs between Western and Eastern cultures. For example, the positivity effect postulates that as people age, they are drawn more to positive stimuli (e.g., happy faces) than to negative stimuli (Isaacowitz et al., 2006). Fung et al.'s (2008) cross-cultural study used eye-tracking technology to analyse participants' gaze behaviour when presented with positive, negative, and neutral stimuli. They found that the positivity effect was only valid for Western cultures (e.g., United States of America). In contrast, older Chinese adults do not exhibit this positivity bias. Instead, they tend to look away from positive stimuli rather than selectively focusing on it. Fung suggests this is because the Chinese culture is more interdependent and, therefore, more sensitive to other people's negative emotions. Furthermore, culture and language are heavily intertwined (Cheang & Pell, 2011; Chen & Dewaele, 2021). For example, even in countries where the language is shared, such as

England and America, humour can be understood differently. Chen and Dewaele (2021) found that Americans often were unable to detect British individuals' subtle ironic sense of humour and that Americans themselves favoured more explicit humour with exaggerated elements. Together, this research suggests that language, culture and mentalising may be interlinked and differences may be found across cultures. Therefore, it is possible that correlations found across different cultures between individuals' differences in mentalising and verbal abilities may be confounded by cultural differences.

A further notable finding of this review is that a small number of tasks have been consistently used to measure the relationship between mentalising and language since 2002. Specifically, the following seven tasks represented 68% of the 90 mentalising tasks used across 182 studies: the Reading the Mind in the Eyes Test, the Strange Stories task, the Faux Pas Task, the Awareness of Social Inference Test, the Hinting Task, the Frith-Happé Animations Test, and the Movie for the Assessment of Social Cognition. These findings align with those of Osterhaus and Bosacki's (2022) systematic review, in which the authors found that five tasks (the Strange Stories task, the Reading the Mind in the Eyes Test, false belief reasoning, the Faux Pas Task, and the Frith-Happé Animations Test) accounted for 63% of all tests used. One difference between Osterhaus's review and the current one is that the most frequently used task in their study was the Strange Stories task (20%), whereas in this study it was the Reading the Mind in the Eyes Test (30%). These different results may be attributed to the fact that the present review targeted studies that have explicitly examined the relationship between mentalising and language. Researchers may have used the Reading the Mind in the Eyes Test more frequently than other tasks to reduce the number of shared task demands between the mentalising and language tasks, which may also explain the more frequent use of film-based mentalising studies – e.g., those employing the Awareness of Social Inference Test (5%) and the Movie for the Assessment of Social Cognition (3%) – in

this review. However, verbal tasks were also commonly used to examine the relationship between mentalising and language, such as the Strange Stories task (11%), the Faux Pas Task (9%), and the Hinting Task (5%).

An important point highlighted by this review is that a wide range of mentalising tasks, which target different subcomponents of mentalising and use different formats, are available to researchers. Across 182 studies, this review identified 91 original mentalising tasks, 20 of which have been incorporated into the taxonomy. The taxonomy was designed to describe the mentalising subcomponents that each task measures and some of the non-mentalising skills recruited in completing these tasks. The taxonomy outlines the different formats that have been used to examine different conceptual categories of mentalising (e.g., higher-order false beliefs). In addition, many of these tasks call on multiple mentalising concepts, and the taxonomy outlines which concepts each task draws upon. For example, the taxonomy shows that higher-order false beliefs have been operationalised using several formats (still cartoons, vignettes, and film). One of these tasks is the Visual Cartoon Task (Corcoran et al., 1997), which draws upon false beliefs as well as cognitive mentalising, lies, and jokes. The Visual Cartoon Task also has control conditions that examine participants' knowledge of physical causality. All these details are described in the taxonomy and should enable researchers to make informed decisions when using mentalising tasks in their studies.

3.5.2 Using the taxonomy to examine the relationship between mentalising and language

To reduce the number of shared task demands when examining the relationship between mentalising and language, it is important to describe the mentalising and verbal

properties of the tasks, as has been attempted in the taxonomy. For instance, the Strange Stories task has been used to show links between mentalising and language in adolescents (Boerma et al., 2017; Ricketts et al., 2013). However, it is a text-based task and contains verbal task demands that researchers might want to avoid. The taxonomy shows that the Strange Stories task targets the mentalising concepts of false beliefs, higher-order intentionality, desires, and pragmatic language (irony, misunderstanding, figure of speech, and persuasion) (White et al., 2009). Researchers could then pair the Strange Stories task with a task that tests the same concepts but uses a different format. For example, the Adult-Mentalising Task (A-ToM; Brewer et al., 2017) is a film-based task which contains scenes based on the Strange Stories task and thus covers the same mentalising concepts. If a researcher decided that the A-ToM was unsuitable for a project or could not obtain this task, then they could use two different tasks to cover the same concepts. A task that covers all the same mentalising concepts as the Strange Stories task, apart from false beliefs, is the film-based Versailles-Situational Intentional Reading (V-SIR), which tests higher-order intentionality, lies, hints, and indirect speech (Bazin et al., 2009; Rambeau et al., 2022). If a researcher deemed false beliefs to be a crucial element of the Strange Stories task that was not covered by V-SIR, then a false belief reasoning task that uses still cartoons, such as the Visual Cartoon Task, could be paired with V-SIR to cover this (Corcoran et al., 1997).

Given the multidimensional nature of mentalising, it would be advisable for future studies to specify which mentalising concepts they are studying rather than stating that they are measuring ‘mentalising’ generally. By referring to taxonomies such as the one produced in this study, researchers should have the flexibility to ‘mix and match’ tasks to form comprehensive multi-component batteries of mentalising tasks that take account of non-mentalising properties recruited. In addition, given that robust associations were found between all the mentalising tasks reviewed (see Section 3.4.6) and vocabulary, any study

interested in the relationship between mentalising and complex language (e.g., reading comprehension and pragmatic language) should control for vocabulary knowledge, and/or include an appropriate higher-order language control. Correlations between mentalising and grammatical knowledge were not frequently reported in the studies identified in this review; however, given that an association between these two constructs has been reported in studies with children (De Villiers, 2000; Fisher et al., 2005), individual differences in grammar may also be important to control for in studies with adults.

2.5.3 Conclusion

There are several important findings in this systematic review. The first is that mentalising has been operationalised in numerous ways in studies of the relationship between mentalising and language in adults. Moreover, many studies use different tasks interchangeably as measures of mentalising, even though they measure different subcomponents. Researchers should use taxonomies such as the one developed in this review to establish which subcomponents of mentalising each task measures. In addition, this review demonstrates that mentalising tasks are presented in diverse formats (e.g., vignettes, photographs, and films). As such, performance on these tasks will recruit different non-mentalising abilities, which should be considered when examining relationships between mentalising and associated abilities. Therefore, to establish whether these links between mentalising and language abilities (including reading comprehension) are due to the mentalising or non-mentalising properties of the mentalising tasks, suitable controls should be put in place (e.g., vocabulary knowledge). Moreover, multiple measures of mentalising should be used to compare which subcomponents of mentalising are related to

other abilities, such as reading comprehension, and to account for non-mentalising task demands.

A key finding from this review is that the same five mentalising tasks are often used to investigate the relationship between mentalising and language:: the Reading the Mind in the Eyes Test, the Strange Stories task, the Faux Pas Task, the Awareness of Social Inference Test, the Hinting Task, the Frith-Happé Animations Test, and the Movie for the Assessment of Social Cognition represented 68% of measurement instances across 182 studies reviewed. However, in this review, 91 original tasks were collated, testing a range of mentalising components in various task formats. The purpose of the taxonomy was to provide researchers with an easy guide to the properties of available mentalising tasks. Detailed examination of the task properties revealed that several tasks do not fit into a single operational category as easily as indicated in Osterhaus and Bosacki's (2022) review. For example, the Strange Stories task was labelled as a 'task that draws on pragmatic language'; however, the present review showed that it correlated less strongly with pragmatic language tasks than with other reading abilities or associated skills (e.g., vocabulary). Further, the Strange Stories task measures a range of different mentalising skills (e.g., affective mentalising and false beliefs), which are described in the taxonomy. This level of nuanced detail in understanding what is measured by different mentalising tasks is important in improving the rigour of research in this area. However, one of the most striking findings from this review is that a widespread understanding of what mentalising tasks measure is lacking, or at least not precisely reported in much of the existing research literature.

4. When Do Comprehenders Mentalise for Pragmatic Inference? a Partial Replication Study

4.1 Chapter overview

Study 1 established associations between mentalising (as indexed by the Strange Stories task) and reading comprehension in adults. The current study (Study 3) explores the mechanisms behind this relationship in more detail. More specifically, this study investigates whether one way in which mentalising may support readers' understanding of a text is by aiding their comprehension of pragmatic language during reading. Spoken dialogue that is used in social contexts (within both oral and written modalities) is referred to as pragmatic language. Interpreting pragmatic language is difficult, because the literal meaning of a sentence may not be sufficient to infer the speaker's intended meaning (Sperber & Wilson, 2002). Therefore, the comprehender must use contextual cues, such as mental state information, to interpret the speaker's intended meaning, which requires mentalising. Hulme and Snowling (2014) note that pragmatic language skills may contribute to the oral language component of the Simple View of Reading. In support of this, Study 2

(systematic review) found evidence suggesting that mentalising has a role in comprehending written pragmatic language, such as in interpreting irony, metaphors, and indirect requests (Ahmed & Miller, 2013; Bambini et al., 2021; Burgers et al., 2018). Hence, Study 3 aims to investigate the role of mentalising in interpreting written pragmatic language.

4.1.1 Introduction

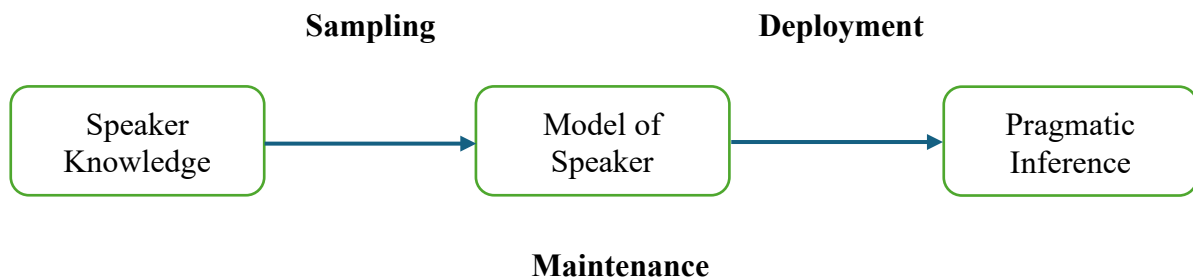
Understanding the words in an utterance alone is often not enough to interpret a speaker's true meaning because people frequently speak indirectly (Sperber & Wilson, 2002). To interpret the meaning behind a non-literal utterance, we need to make a pragmatic inference. To do this, a co-communicator must be sensitive to contextual cues, such as the speaker's knowledge of a given situation (Apperly, 2018; Franco & Galvis, 2013). For instance, if a colleague came to one's office to help with a project and said, 'It's hot in here', this could be a descriptive comment, or alternatively a request for the listener to turn down the central heating. However, contextual cues are often not as simple as this, and so interpreting the true meaning of indirect requests becomes more complicated. It may be necessary to refer to what a person knows to understand their intentions (Gibbs, 1987). For example, suppose the colleague knew the central heating was managed automatically by the building's communal heating system. In that case, the listener might be more likely to interpret their utterance 'It's hot in here' as a descriptive comment. Thus, building a mental representation of what a co-communicator knows and believes may aid in the interpretation of pragmatic utterances. It has been suggested that mentalising, the ability to reason about the mental states of others, may support this process (Sperber & Wilson, 2002). The current study investigates the role of mentalising in text-based dialogues between characters. Although most research in this area has examined perspective-taking in spoken dialogue, a

recent study by Trott and Bergen (2020) used a model by Apperly (2018) to explain how mentalising may support pragmatic inferences during reading.

Apperly's (2018) component model of perspective-taking was originally designed to understand how mentalising may support pragmatic inferences about the meaning of spoken utterances. He suggests that to make a pragmatic inference about the meaning of an utterance (e.g., 'it's hot in here'), individuals must attend to their co-communicator's perspective (e.g., what they see, hear, know). Often, attention cuing and memory retrieval are sufficient to coordinate dialogue between two co-communicators without the need to represent a co-communicator's perspective (Horton & Grieg, 2005; Long et al., 2018; Ryskin et al., 2015; Wardlow, 2013); Apperly (2018) describes this as *perspective sensitivity*. At other times, when two co-communicators have conflicting perspectives (e.g., the listening colleague knows they have no control over the office's central heating while the speaker does not know this), mentalising may be recruited so that a co-communicator's perspective can be fully represented; this is commonly referred to as *perspective representation*. Apperly (2018) uses both perspective sensitivity and perspective representation in his component model of perspective-taking (see Figure 4.1). Here perspectives are encoded over three phases: sampling, maintenance, and deployment. A comprehender must adopt each of these different component processes to interpret the true meaning of a speaker's indirect request. First, the comprehender must *sample* information about a speaker's perspective. Secondly, they must maintain this information in their memory. Thirdly, they *deploy* this information to make a pragmatic inference.

Figure 4. 1

Apperly's (2018) Component Model of Perspective-Taking



Note: This model describes the process of adopting a speaker's perspectives to, for example, interpret the meaning of an indirect request.

Apperly (2018) proposes that mentalising may be employed during each of the three stages of his model to support individuals' pragmatic inferences. In the sampling stage, the co-communicator must build a mental representation of the speaker's perspectives, such as their thoughts and attitudes; this may be done by being attentive to what the speaker can see and hear, as well as by sampling information from their spoken dialogue. During the conversation, Apperly proposes that mentalising may support the generation of inferences about the speaker's perspectives by combining samples of spoken dialogue with the listener's understanding of mental states. In this way, mentalising could be viewed as a particular type of background knowledge. Mentalising may support mental state inference generation in the same way that background knowledge supports the generation of regular inferences during reading by going beyond the text to make sense of the discourse (Anderson & Pearson, 1984). Like regular inferences, mental state inferences are made spontaneously rather than automatically. Back and Apperly (2010) asked participants to watch picture sequences describing people's true or false beliefs regarding the location of an

object. They observed that participants interpreted reality questions (where was the object?) much faster than belief questions (where did they think the object is?), suggesting that the participants were not automatically ascribing the people's beliefs. In addition, the amount of information spontaneously inferred during sampling depends on individual differences in cognitive resources and the listener's motivation to understand the speaker (Surtees et al., 2016).

After information about the speaker's perspectives is encoded during the sampling stage, this mental state information must be maintained in memory. At the maintenance stage, it has been shown that holding information about beliefs that contradict one's own can cause interference with cognitive processing. Mentalising may be recruited at this stage to aid the representation of these contrasting perspectives in memory (Back & Apperly, 2010; Samson et al., 2010). In one study, Apperly et al. (2008) gave participants written information about a character's belief about an object (e.g., 'He thinks the ball on the table is red, but really the ball on the table is blue') and then presented them with a picture of a scene. Participants were slower at responding to questions about the character's belief when the character's belief was false than when they had a true belief about the location of the object. This indicates that even when mental state information is sampled correctly, interference in memory is caused by the conflict between the individual's perception of the scene (true belief) and the character's conflicting perspective (false belief). However, the participants still needed to access and deploy the information in their memory in order to say whether the picture they saw matched the statement they previously read. Thus, this experiment may measure maintenance and deployment. Although it cannot be denied that people must hold perspective information in memory in order to use it at a later point, the maintenance stage is the least studied stage of Apperly's component model; isolating maintenance from deployment in experimental design is challenging.

Finally, in the deployment stage, the listener may need to represent mental state information to accurately interpret the intentions behind pragmatic inference. At this stage, individuals need to draw on the information encoded during sampling and stored during maintenance and deploy it to make a pragmatic inference about the meaning of an utterance. For example, when interpreting a friend's ironic comment regarding an event (e.g., 'Well, that was time well spent!'), we need to represent the speaker's true feelings (e.g., they felt the event was a waste of time) to understand that the literal meaning of the utterance did not represent the intended meaning of their utterance (Keysar, 2000). Individual differences in deployment for pragmatic inference in adults have been traced using the Director's Task (Dumontheil et al., 2010; Keysar, 2003), which taps visuospatial perspective taking (see Section 2.3.1 for a full description). In this task, the participant must sample and maintain information about their own and the director's different perspectives of objects in a grid. Then the participant must deploy this information when following the director's verbal instructions to move the objects around the grid. Participants from clinical groups associated with differences in social communication (e.g., autism, depression) produce lower scores in the Director's Task than controls, indicating that it is a valid measure of social communication (Abu-Akel et al., 2015; Nilsen & Duong, 2013). Furthermore, research examining children's performance on the Director's Task indicates that individual differences in working memory and language comprehension skills explain the variance in performance during the deployment phase of this task; this suggests that deployment draws on working memory and language comprehension abilities.

4.1.2 Pragmatic language in written text

When interpreting pragmatic language, a key problem for listeners (or readers) is that the literal meaning of an utterance is often not enough to interpret the speaker's intended meaning. The listener must also integrate information from contextual cues with what is said (Sperber & Wilson, 2002). Contextual cues may include the preceding dialogue, knowledge of events surrounding the conversation, general knowledge, and knowledge of the speaker's intentions or perspectives (Frank, 2018). Pragmatics can include interpreting irony, metaphors, and indirect requests (Burgers et al., 2018). As the Gricean view (Grice, 1989) regards pragmatic inferencing as reliant on attributing communicative intentions to the speaker, post-Griceans argue that pragmatic inferencing should be associated with mentalising. Sperber and Wilson (2002) describe pragmatic language as a subcomponent of mentalising with its own unique mechanisms. However, more recent research suggests that, while pragmatic inferencing relies on representing others' perspectives, it also relies on other cognitive functions, such as working memory and attention (Apperly, 2018; Warnell & Redcay, 2019). Therefore, Apperly's component model of perspective-taking may help understand how mentalising supports pragmatic language interpretation during reading.

Although Apperly's (2018) model was based on spoken dialogue between co-communicators, it has been adapted for use with dialogue embedded in written text by Trott and Bergen (2020). Interpreting pragmatic language within a narrative story presents the reader with obstacles that they would not encounter during spoken dialogue, because many of the cues one would use to disambiguate an indirect utterance are no longer available. When reading, we do not have access to visual cues, such as facial expressions or gestures, or auditory cues, such as prosody. Moreover, our general knowledge of past encounters between characters is restricted to the information the author provides (Tannen & Wallat,

1987). As such, readers must be more attentive to key pieces of information that may aid the interpretation of pragmatic utterances, placing greater strain on attention or memory resources. The reduced number of contextual cues makes it harder to interpret the speaker's intentions and to disambiguate non-literal utterances. As a result, there may be a greater reliance on mentalising ability in interpreting dialogue in written text than in spoken communication.

Apperly speculates that in each stage of perspective taking (sampling, maintenance, and deployment), the co-communicator's perspectives may be represented or may be coordinated by perspective sensitivity alone. Trott and Bergen (2020) built on this model by examining when if and when mentalising is recruited during the interpretation of pragmatic utterances in written text. Trott and Bergen (2020) speculate that mentalising might be differently recruited during Apperly's three stages. Mentalising may be primarily recruited during sampling and maintenance, in order to build a coherent mental representation of what a speaker knows or believes. Deploying mental state information to make a pragmatic inference may simply rely on the comprehender's ability to manipulate this information in working memory. Alternatively, sampling and maintenance may be coordinated via perspective sensitivity processes, and mentalising may be recruited during deployment to interpret the intended meaning of this information. For example, when interpreting a sarcastic comment, an individual may correctly sample what a person said and maintain it in memory but may need to reason about the speaker's mental state to infer whether the comment was meant sarcastically or not (Keysar, 2000). To explore these processes, Trott and Bergen (2020) conflate sampling and maintenance, since there is no clear method that disassociates these processes.

Trott and Bergen (2020) experimental design is based on the premise that if individuals adopt the speaker's perspectives to interpret the meaning of an indirect request,

then altering the speaker's knowledge states (what they know or believe) should systematically change an individual's interpretation of an indirect request. If perspective-taking is indeed a component process, then differing effects would be found during sampling, maintenance, and deployment (see Figure 4.2).

Figure 4. 2

Aware and Unaware Conditions in the Inference Group from Trott and Bergen's (2020)

Stimuli

(AWARE CONDITION) You and your friend Jonathan are taking a road trip. You began in California and are now passing through Michigan. It's almost winter, so it's very cold outside - especially for Southern California dwellers like you and Jonathan. You see that you're almost out of petrol, so you stop at a petrol station in a small town. You fill up the tank, and then the **two of you** go inside the petrol station to buy some water and snacks. When you return to the car and start up the engine, you and Jonathan notice with some dismay a blinking light, which indicates that the car's heating system is broken. You both wrap up warm. As you leave the station, Johnathan shivers in his seat. He turns to you and says, "Man, it's really cold in here."

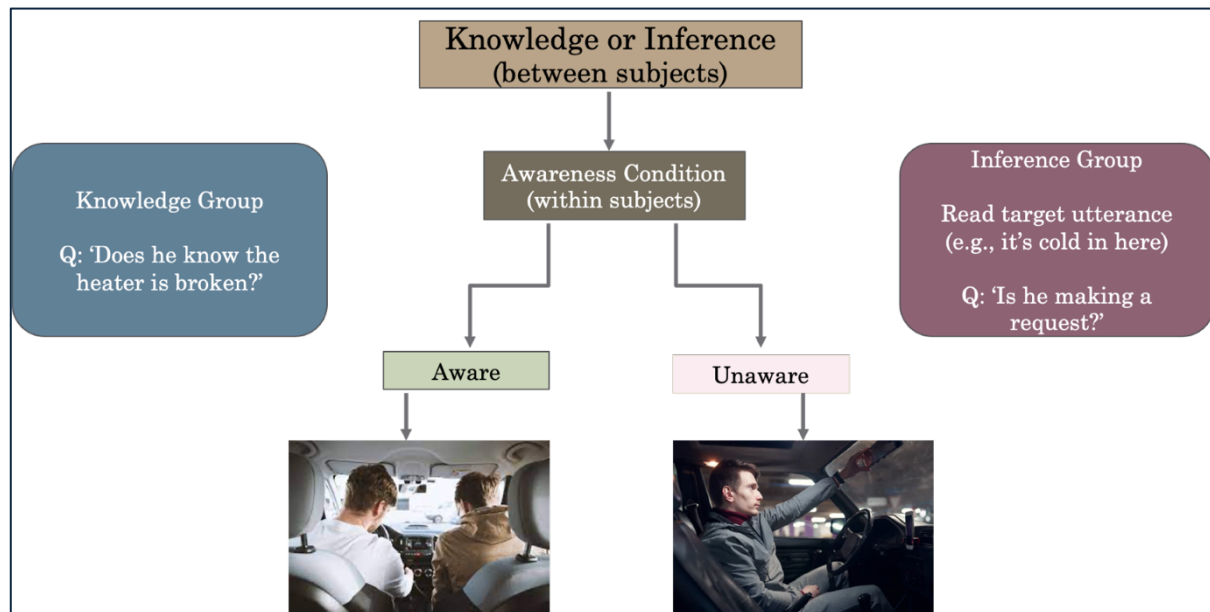
(UNAWARE CONDITION) You and your friend Jonathan are taking a road trip. You began in California, and are now passing through Michigan. It's almost winter, so it's very cold outside - especially for Southern California dwellers like you and Jonathan. You see that you're almost out of petrol, so you stop at a petrol station in a small town. While you fill up the tank, **Jonathan goes inside** to buy some water and snacks. As you're checking the meter, you notice with some dismay a blinking light, which indicates that the car's heating system is broken. You finish filling up the petrol and wait for Jonathan. Jonathan returns with some snacks, and you both set off.

As you leave the station, he shivers in his seat. He turns to you and says, "Man, it's really cold in here."

In Experiment 1 from Trott and Bergen (2020), a mixed factorial design was devised, with one within-subject condition (awareness) and one between-subject condition (group-type). In the awareness condition, participants read eight pairs of narratives that described an interaction between two characters. In each pair of narratives, the protagonist was either aware or unaware of an obstacle to fulfilling a request (e.g., the car's heater is broken). In the between-subject factor, participants were placed in one of two groups: knowledge or inference. After reading the narrative, Knowledge Group participants were asked a knowledge-based question ('Does he know the car's heater is broken?') that targeted sampling and maintenance. In the Inference Group, an additional target utterance was added at the end of the story (e.g., one character says, 'It's cold in here'), and the participants were asked an inference-based question ('Is he making a request?'). These questions targeted deployment (see Figure 4.3).

Figure 4. 3

Trott and Bergen's (2020) Mixed Design Study



Note: Participants read a set of narratives in which the speaker's perspective was manipulated so that the speaker was either aware or unaware of an obstacle to fulfilling a request. In the Knowledge Group, they were asked a question about the speaker's knowledge state. In the Inference Group, the participants read an additional target utterance at the end of the narrative and had to decide whether the speaker was making a request or not.

In each group (knowledge or inference), the participants were assigned to one of two lists of narratives, counterbalanced for aware or unaware versions of the story. The aim was to see if the participants modulated their answers as a function of awareness in both the Knowledge and Inference Groups. The results of this study indicated that manipulating speaker knowledge states modulated individuals' responses during all component stages of perspective-taking, but the effect was greatest during sampling and maintenance. In addition, they found that individual differences in mentalising ability (measured using the

Dodell-Feder et al.'s (2013) Short Story task) predicted sampling and maintenance, but not deployment, of speaker perspectives. These findings indicate that mentalising is important when constructing a mental model of what the speaker knows or believes but that the process of using this information to make a pragmatic inference (deployment) relies on other processes. The current study sought to replicate these findings and to investigate which other processes may be involved in the deployment of knowledge-state information to make pragmatic inferences.

To demonstrate the involvement of mentalising skills in the comprehension of dialogue in written text, it is necessary to replicate these results using additional mentalising measures. Mentalising research has greatly expanded over the last 40 years, and with this expansion has come a range of highly heterogeneous mentalising measures that vary across dimensions (see Chapter 3). Recent mentalising research has questioned the reliability, validity, and coherence of mentalising tasks and suggests that experimental studies should include multiple measures to understand which facet of mentalising is involved in a particular process (Bialecka-Pikul et al., 2021; Warnell & Redcay, 2019). Trott and Bergen (2020) discovered that a text-based measure of mental state reasoning (i.e., the Short Stories task) predicts perspective-taking during reading comprehension. To extend these results, this study examines whether other measures of mental state reasoning predict performance in Trott and Bergen's experimental paradigm. In the present study, two mentalising measures were used in place of Short Stories, given its psychometric weaknesses (see Chapter 3): Strange Stories (White et al., 2009) and the Frith-Happé Animations Test (Campbell et al., 2006; Livingston et al., 2019). Strange Stories is a text-based mentalising task that has been shown to be sensitive to individual differences in the adult population (Happé et al., 1998; Wang & Su, 2006). As there are shared task demands between text-based measures of mentalising and Trott and Bergen's (2020) Pragmatic Inference Task, a non-text-based

measure of mentalising was also included. The Frith-Happé Animations Test measures the ability to attribute human intentions to two interacting shapes (Livingston et al., 2019). In a study by Ricketts et al. (2013), scores on both the Strange Stories and the Frith-Happé Animations predicted reading comprehension scores in autistic adolescents aged 14 to 16 years old. Although Frith-Happé Animations recruits mental state reasoning skills, it is likely to also tap visuospatial perspective-taking. It is therefore possible that the Strange Stories and the Frith-Happé Animations Test measure different facets of mentalising (see Chapter 3). However, using the Frith-Happé Animations Test avoids overlapping reading comprehension skills that may be common to text-based mentalising tasks, such as Short Stories and Strange Stories, and Trott and Bergen's (2020) perspective-taking paradigm.

Trott and Bergen (2020) controlled for reading comprehension because the test of mental state reasoning that they used, namely the Short Story Task, was text-based, but there were no other control variables. Previous research (Horton & Grieg, 2005; Long et al., 2018; Ryskin et al., 2015; Wardlow, 2013) suggests that perspective-taking is coordinated by perspective sensitivity processes, such as working memory. It is possible that when working memory is controlled, the relationship between mentalising and the sampling and maintenance components of perspective-taking will be reduced or eliminated. This idea is consistent with recent research in healthy ageing that analysed a variety of pragmatic language tasks and concluded that working memory predicted performance over other cognitive skills, demographic features, and mentalising (Bambini et al., 2021).

Trott and Bergen's (2020) study gives unique insights into how mentalising may be involved in reading comprehension, and its methodology can be used to explore the role of other sources of individual variation associated with mentalising and reading comprehension. Lifespan studies have shown that adults' performance on tests of mental state reasoning only starts to decline after 60 years old (Cavallini et al., 2013), whereas other

research shows that mentalising is relatively well preserved up until 80 years old (Giovagnoli, 2019). It may be that mentalising continues to support the creation of rich mental representations of a text during narrative text comprehension in later adulthood when skills such as working memory have begun to decline (Wang et al., 2011). This could explain why there is a decline in expository text comprehension, whereas narrative text comprehension is relatively preserved in older adulthood (Artuso & Bellacchi, 2021; De Beni et al., 2007). In expository texts, characters' motivations are often less intrinsically entwined with the events in the text, and so mentalising ability is less likely to be recruited in reading comprehension. Study 1 found that age positively correlated with mental state reasoning, reading comprehension measures, and vocabulary in adults aged 18–68, so age effects might be expected in Trott and Bergen's (2020) Pragmatic Inference Task.

Present Study

Study 3 involved a partial replication and extension of Trott and Bergen's (2020) Experiment 1. The first part of the study is a pure replication that examined (1) whether manipulating the speaker's knowledge states systematically altered adult participants' ability to sample and maintain speaker information about what the speaker knows or believes and (2) their ability to deploy that information to make a pragmatic inference about the meaning of an indirect request. In line with Trott and Bergen's (2020) results, it was predicted that in both groups (knowledge and inference), participants would modulate their responses as a function of the awareness condition but that they would modulate their responses to a greater extent in the Knowledge Group. A second aim of this study was to partially replicate the second stage of Trott and Bergen's (2020) study by (1) using different mentalising measures to predict performance on the pragmatic inference task and (2) including working

memory and age, in addition to reading comprehension, as control variables. Assuming that mentalising has a unique role in pragmatic inference generation during reading, it was hypothesised that better mentalisers would be more likely to choose the correct result in the Knowledge Group when participant age, working memory and reading comprehension are controlled.

4.2 Method

Participants

Participants were recruited online via two platforms. The first was Prolific, and the second was SONA, the York St. John University online recruitment service. Prolific participants were paid £7.50 for participation, and SONA participants were given course credits. There were four sets of pre-screening criteria: (1) age (participants had to be over 18), (2) nationality (participants had to come from the United Kingdom or Ireland), (3) first language (participants had to speak English as a first language), (4) no literacy difficulties.

As this study aimed to examine mentalising throughout adulthood (18–60 years old), recruitment took place in three phases to obtain a roughly equal distribution of participants across this age range (see Table 4.1). Sixty participants were recruited via SONA, and 109 participants were recruited via Prolific. Two participants were rejected from the Prolific sample for repeatedly failing to make a reasonable attempt to answer the questions. Therefore, the total number of participants was 207.

Table 4. 1*Demographic Statistics*

Recruitment Waves	Number of participants	Age Mean (SD)	Gender
18–28	70	20.41 (2.74)	63 females, 6 males, 1 other
29–44	70	35.34 (4.44)	42 females, 28 males
45–60	67	52.87 (3.90)	27 female, 40 male
Whole Sample	207	35.97 (13.75)	132 females, 74 males, 1 other

*Materials**Pragmatic Inference Task (Experimental Task; Trott & Bergen, 2020)*

The Pragmatic Inference Task examines how manipulating speaker knowledge (e.g., what the speaker knows or does not know) affects participants' interpretation of an indirect request. In this task, there were eight pairs of narratives that described an interaction between the protagonist (the speaker) and a second character (e.g., they drive to a garage to fill up the petrol on a cold night). During the narratives, the second character learned about an obstacle to fulfilling a request (e.g., the car heater was broken). Within the pairs of narratives, the speaker's knowledge of the obstacle was manipulated, and the speaker was either aware or unaware of the obstacle. Participants saw either an aware or unaware version of the narrative (see Figure 4.4).

Figure 4. 4

Trott and Bergen's (2020) Pragmatic Language Task (Study 2)

Speaker Aware

You and your friend Jonathan are taking a road trip. You began in California and are now passing through Michigan. It's almost winter, so it's very cold outside – especially for Southern California dwellers like you and Jonathan. You see that you're almost out of gas, so you stop at a gas station in a small town.

You fill up the tank, and then the two of you go inside the gas station to buy some water and snacks. When you return to the car and start up the engine, you and Jonathan notice with some dismay a blinking light which indicates that the car's heating system is broken. You both bundle up.

As you leave the station, Jonathan shivers in his seat. (He turns to you and says, 'Man, it's really cold in here.')

Prompt for Sampling and Maintenance Group *Does he know the heater is broken?* **Prompt for**

Pragmatic Inference Group *Do you think he is making a request?*

Note: This is an example of an aware condition narrative with the items (prompts) below. In the unaware condition, Jonathan goes into the gas station and does not see the heater is broken. Participants were assigned to the Sampling and Maintenance Group or the Inference Group, and the number of aware and unaware versions of the story was counterbalanced within each group. The text in parentheses for items 1– 8 represents the target utterance for the Inference Group.

Participants were split into two groups: the Knowledge Group and the Inference Group. In the Knowledge Group, participants read the narratives, followed by a question about the speaker's knowledge states (e.g., 'Does he know the heater is broken?'). In the

Inference Group, an additional utterance by the speaker is added at the end of the narrative (e.g., ‘It’s cold in here’) that is not present in the Knowledge Group narratives. This is followed by an inference question (e.g., ‘Is he making a request?’). In both groups, all questions were answered with a ‘yes’ or ‘no’. Each question item was scored 1 or 0 (correct or incorrect). A maximum score of 8 was possible (1 point for each narrative).

Mentalising measures Attributing Intentions to Shapes Task (Frith-Happé Animations)

In the Frith-Happé Animations Test (Campbell et al., 2006; Livingston et al., 2019) the participant sees eight animations (four ‘physical interactions’ and four ‘mental interactions’) in which triangles interact with each other. The interaction between the triangles follows a script designed by the original authors (Abel et al., 2000) in which the triangles enact a scene. In the goal-directed animations, the triangles react to one another’s physical movements (e.g., fighting, dancing), whereas in the mentalising animations, the triangles react to each other’s mental states (e.g., surprised, seducing). In the practice round, the participant saw two animations, one goal-directed and one mentalising, and after guessing whether they animations depicted a goal direct or mentalising interaction, they were given the correct answer. After the participants watched the animations (each 40 seconds in length, approximately), they were asked to write a description of the interaction between the triangles in a textbox. There were four goal-directed and four mentalising animations in the experimental rounds.

There were two stages to the scoring. In the ‘objective scoring’ stage, a total of 8 points could be awarded for correctly identifying whether it is a physical or mental interaction (1 point per animation). In the ‘subjective scoring’ stage (Campbell et al., 2006), the participant’s description of the interaction between the triangles was coded using two different scoring systems. To calculate the intentionality score, the participant was awarded

points for the intentional language they used (e.g., the big triangle *wanted* to find the little triangle; scored on a rating scale of 0–5 for each animation). For the appropriateness score, the participant was awarded points according to how appropriately they describe the scripted interaction between the triangles (0–2 for each animation). The intentionality scores and appropriateness scores were added together to give a total score of 28 each for the four goal-directed and four mentalising animations. The Frith-Happé ToM composite score was produced by combining the Frith-Happé appropriateness and intentionality scores, and this was used as the final score on the Frith-Happé animations. The intentionality scores had a Cronbach alpha of .67, and the appropriateness scores had a Cronbach alpha of .57.

Inter-rater reliability in the Frith-Happé Animations

The Frith-Happé Animations appropriateness and intentionality scores were coded by the main researcher and two other coders. This initial coding phase ensured that the coders shared conceptual interpretations of each item. In the initial session, the data from 20 participants were coded, and all interclass correlations (ICCs) were above .70. Discrepancies were then discussed and resolved in a meeting between the coders. Two coders then coded the remaining 187 participants' data. In this second phase of coding, a high degree of reliability was found between the two coders using intraclass correlations. The intentionality scores had an intraclass correlation of .83, with a 95% confidence interval of .79 to .87 ($p < .001$). The intraclass correlation of the appropriateness scores was .90, with a 95% confidence interval of .88 to .93 ($p < .001$). Any discrepancies were then discussed and agreed upon by the two coders.

Excluded data

After reviewing the Frith-Happé Animations Test, data from 24 participants were excluded. When asked to write a description of the interaction between the triangles, 21 participants gave an insufficient answer and simply wrote ‘mental’ or ‘physical’. This was likely due to misunderstanding the task instructions. Data for three other participants were removed due to technical errors, and one person repeatedly wrote that they did not understand the task. This left a total of 183 datasets.

Mental State Reasoning (The Strange Stories Task)

A short-form version of Strange Stories (White et al., 2009) was used as a measure of mental state reasoning (see Section 2.2.2 for a full description of this task). The eight vignettes used in the current study required an understanding of white lies, double bluffs, persuasion, and misunderstanding. After reading each vignette silently, the participant was asked a single mental state question about the protagonist’s behaviour (White et al., 2009). Responses were given verbally and recorded. Accuracy was rated on a 0–2 scale by two researchers: 0 points for an incorrect answer, 1 point for a partially correct answer, and 2 points for a fully correct answer. The maximum score was therefore 16 points. The Strange Stories had a Cronbach alpha of .52.

Reading Comprehension

A short-form version of the Strange Stories Physical Stories (Strange Stories Physical; White et al., 2009) was used as a measure of reading comprehension. The Strange Stories Physical were designed as a control measure to be used alongside the Strange Stories Mental State Reasoning; the task demands in the Strange Stories Physical are exactly the

same as those in the Strange Stories' mental state reasoning except that in these stories, the individuals must reason about physical states only, which means this is an effective control measure. After reading each vignette silently, the participant is asked a single mental state question about something the protagonist said or did. Responses were given verbally and recorded. Accuracy was rated on a 0–2 scale by two researchers: 0 points for an incorrect answer, 1 point for a partially correct answer, and 2 points for a fully correct answer. The maximum score was 16 points. The Strange Stories Physical had a Cronbach alpha of .63.

Inter-rater reliability in the Strange Stories and Strange Stories Physical Tasks

The Strange Stories and the Strange Stories Physical were also coded by the lead researcher and two other coders. All coders rated 20 scripts. Then, once inter-rater reliability was found to be acceptable, the remaining 187 scripts were divided between the coders and coded. A high degree of reliability was found between the coders; all ICCs were above .70, indicating a good level of reliability. For the Strange Stories, intraclass correlation was .80, with a 95% confidence interval of .73 to .85 ($p < .001$). For Strange Stories Physical, the intraclass correlation was .86, with a 95% confidence interval of .81 to .91 ($p < .001$).

Working Memory

To test working memory, a forwards and backwards Digit Span Task was used (Turner & Ridsdale, 2001). In the forward-span task, the participants listen to a string of numbers and must repeat them back in the order in which they heard them. After each round, the string of numbers to remember increases by one digit. In the backwards-span task, the participants listen to a string of numbers and must repeat them back in the opposite order in which they heard them. After each round, the string of numbers to remember

increases by one digit. The digit span forward and backwards each had nine rounds. For example, in round one the participant had to remember two digits, and in the last round they had to remember 10 digits. The individual's digit span score is the maximum length of the lists where the participant recalled both the forward and backward tasks. The Digit Span Task had a Cronbach alpha of .80.

Procedure

Participants provided informed consent prior to participating in the experiment. All testing took place online using the software platform Gorilla, and testing took approximately 45 minutes to complete. To begin, participants were randomly allocated to either the Knowledge Group (103 participants) or the Inference Group (104 participants) using Gorilla's randomisation and quota nodes. Participants were then directed to one of two lists counterbalanced for 'Speaker Aware' and 'Speaker Unaware' versions of the story. Participants in both groups then read eight narratives; at the end of each narrative, they advanced to the next page and answered a knowledge-based (e.g., 'Does he know the heater is broken?'), or inference-based (e.g., 'Is he making a request?') question by clicking a 'yes' or 'no' button.

After completing the Pragmatic Inference Task, all participants completed all mentalising and control tasks in the same order: the Digit Span, Strange Stories and Strange Stories Physical, and finally the Frith-Happé animations.

Data Analysis

The statistical software R (R Core Team, 2021) was used to process all analyses. Generalised linear mixed-effects models were run using the `glmer` command from the `lme4` package (Bates et al., 2015). In the analysis of both groups, the full model included the

response ('Yes' or 'No') as the dependent variable, the Awareness condition (Speaker Aware or Speaker Unaware) as the fixed effect, and random intercepts and slopes, both by-subject and by-item, for the effect of Awareness. Nested model comparisons were used to obtain results, and the Likelihood Ratio Test was run to test for significant differences between the models. To identify the random effect structure, a maximal model was built with random intercepts and random slopes for by-subject and by-item random effects. Model convergence was then obtained by reducing random effects where needed (Barr et al., 2013; Brauer & Curtin, 2018). The study was registered with the [Open Science Framework](#).

4.3 Results

Table 4. 2

Descriptive Statistics

Task	Mean (SD)	Range	Max. Score
Strange Stories Mental	12.97 (2.11)	7–16	16
Strange Stories Physical	10.50 (2.65)	3–16	16
Digit Span	5.49 (1.66)	2–9	10
<i>Frith-Happé -ToM</i>			
Appropriateness	3.27 (2.06)	0–8	8
Intentionality	12.26 (3.07)	3–19	20
ToM composite	15.53 (4.69)	3–27	28
<i>Frith-Happé - GD</i>			
Appropriateness	4.99 (2.27)	0–8	8
Intentionality	8.05 (3.40)	0–14	20
GD composite	14.82 (2.35)	8–20	28

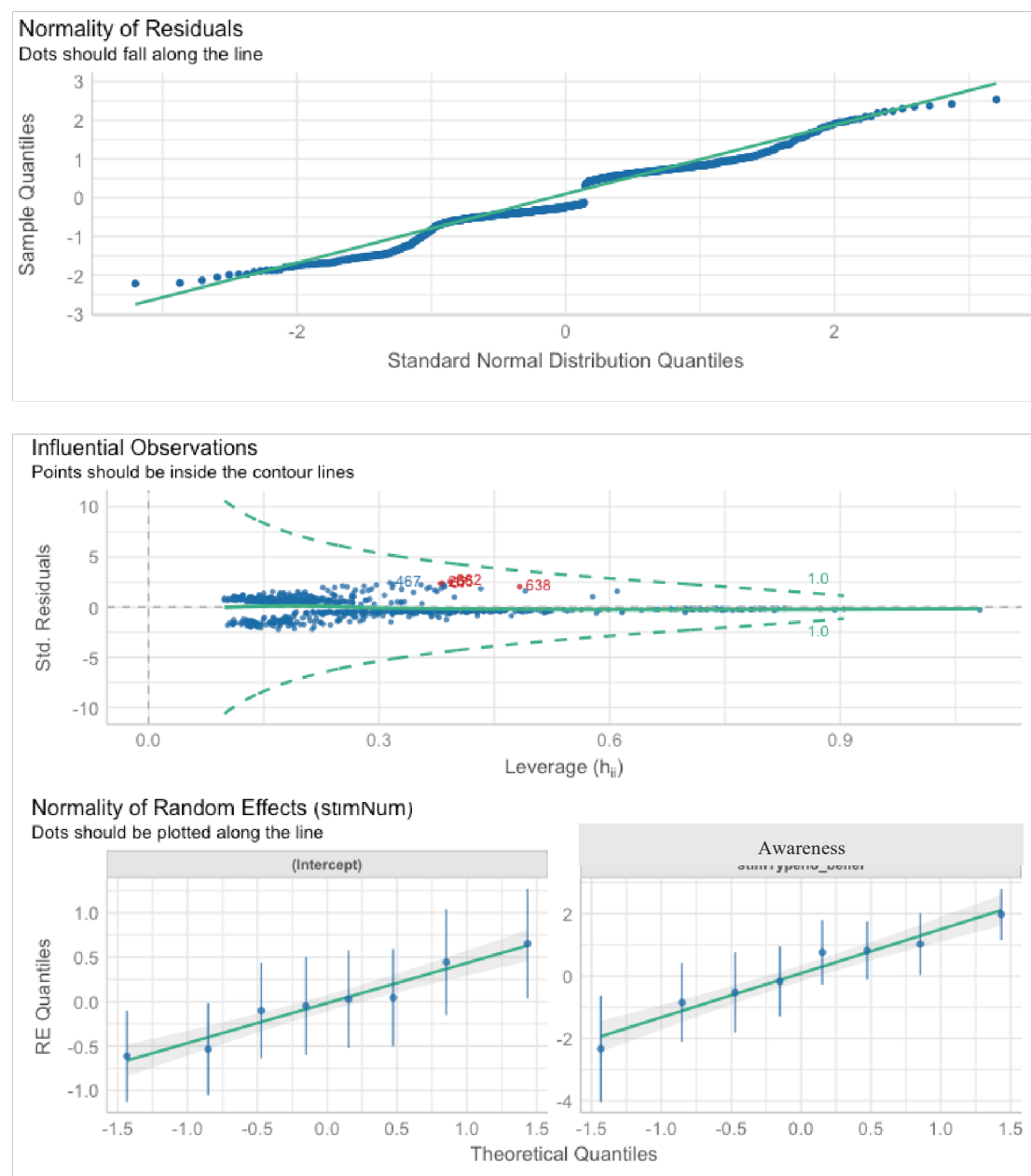
Note. Frith Happé – ToM = Frith-Happé mentalising animations, Frith-Happé GD = Frith-Happé goal directed animations. The ToM composite score was produced by combining the Frith-Happé appropriateness and intentionality scores, and this was used as the final score on the Frith-Happé animations.

Descriptive statistics for all measures are presented in Table 4.2. The Strange Stories did not have ceiling effects and showed variation between participants, indicating it was a suitable measure for adults. To analyse the model's assumptions, the ‘check_model()’ function from the performance package was run. This function always checks for homogeneity of variance, multicollinearity, influential observations (outliers), and normality

of residuals (see Figure 4.5). Evaluating the assumptions, evidence of multicollinearity was found, but this was removed when the variables were grand mean-centred (see below). All other assumptions were satisfied.

Figure 4. 5

Figures Show the Checks for Homogeneity of Variance, Multicollinearity, Influential Observations (Outliers), and Normality of Residuals



The correlations between the mentalising variables and the control variables are shown in Table 4.3. There were small correlations between the mentalising variables, Strange Stories and the Frith-Happé Animations ($r = .19, p < .001$). There were moderate correlations found between the mentalising variables and the control tasks. The correlation between the Frith-Happé Animations mentalising and Frith-Happé Goal-Directed (control) was $.40, p < .001$, and the correlation between the Stranger Stories (mental) and Strange Stories Physical (control) was $.46, p < .001$. There were also significant correlations between the Frith-Happé Animations mentalising and reading comprehension ($.20, p < .001$) and between the Frith-Happé Goal-Directed and reading comprehension ($.22, p < .001$). Finally, a small significant correlation was found between reading comprehension and age ($.21, p < .001$).

Table 4. 3

Correlations Between Mentalising Variables and Controls

	1.	2.	3.	4.	6.	7.
1. Frith-Happé (mentalising)						
2. Frith-Happé (Goal Directed)	.40**					
3. Strange Stories	.19**	.12				
4. Reading comprehension (Strange Stories Physical)	.20**	.22**	.46**			
5. Age	-.08	-.13	.12	.20**		
6. Digit Span	.01	.05	.11	.07	.08	

* $p < .05$, ** $p < .001$

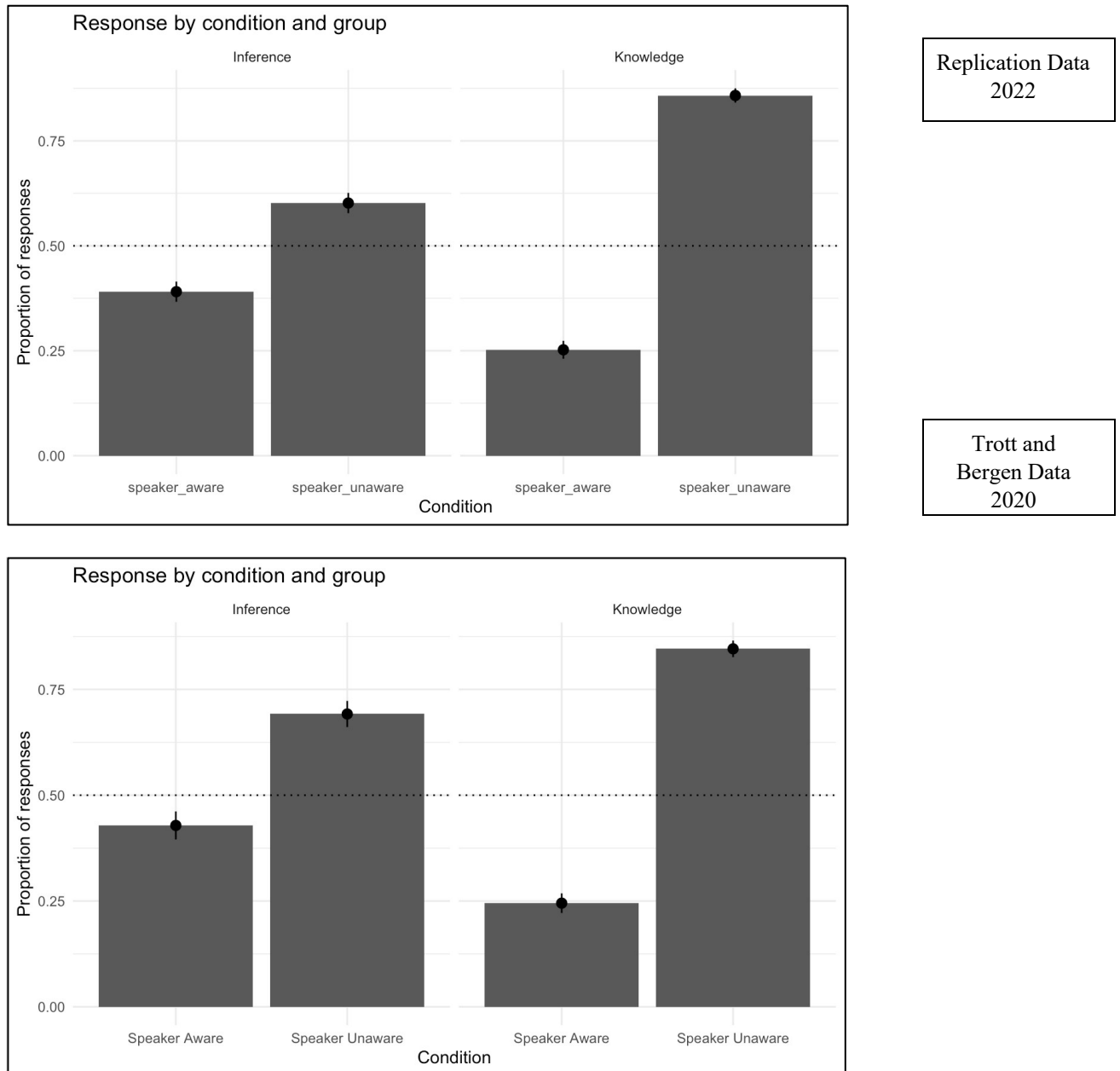
Effects of manipulating speaker knowledge on indirect request comprehension (pure replication)

The first stage of the analysis aimed to replicate the results of Trott and Bergen (2020) to assess whether manipulating speaker knowledge state produced systematic changes in the way that people sample and maintain speaker knowledge states (Knowledge Group) and how they deploy this information to interpret the meaning of an indirect request (Inference Group). Following Trott and Bergen, it was predicted that a significant effect of speaker knowledge (awareness) in both groups (Knowledge and Inference) would be observed. More specifically, in the Knowledge Group, it was expected that a greater number of ‘No’ responses (i.e., ‘No, he is not aware of the obstacle’) would be found in the Speaker Unaware condition. In the Inference Group, it was expected that a greater number of ‘Yes’ responses (i.e., ‘Yes, he is making a request’) would be found in the Speaker Unaware condition.

In the analysis of both groups, the full model included the response (‘Yes’ or ‘No’) as the dependent variable; the awareness condition (Speaker Aware or Speaker Unaware) as the fixed effect; and random intercepts and slopes, both by-subject and by-item, for the effect of awareness. This model was then compared to a null model that did not contain the fixed effect (awareness). It was expected that the model with the fixed effect would explain more variance than the model without the fixed effect.

Figure 4. 6

Comparison of the Results Obtained in this Study with Trott and Bergen's (2020) study



Note. Trott and Bergen's results were replicated

In the Knowledge Group, the results replicated the findings of Trott and Bergen's (2020) study (see Figure 4.7). The full model, including the fixed effect of awareness,

explained significantly more variance [$\chi^2(1) = 23.35, p < .001$] than a model omitting the fixed effect ($p = .138$). As presented in Figure 4.7, there were substantially more significant ‘No’ responses (i.e., ‘No, he is not aware of the obstacle’) in the Speaker Aware condition (86%) compared to the number of ‘No’ responses in the Speaker Unaware condition (25%), [$\beta = -3.51, SE = 0.55, p < .001$]. This is very similar to the proportion of responses found in Trott and Bergen’s (2020) study, where 85% of participants responded ‘No’ in the Speaker Unaware condition compared to 25% of participants in the Speaker Aware condition [$\chi^2(1) = 23.33, p < .001$].

The results also replicated Trott and Bergen’s (2020) findings in the Inference Group and likewise indicated that the effect of awareness was smaller than in the Knowledge Group. The full model including the fixed effect of awareness explained significantly more variance [$\chi^2(1) = 7.67, p = .005$] than a model that omitted the fixed effect ($p = .597$). There were substantially more ‘yes’ (i.e., ‘Yes, he is making a request’) responses in the Speaker Unaware condition (60%) than in the Speaker Aware condition (39%), [$\beta = 1.0, SE = 0.29, p = .001$]; this was also similar to Trott and Bergen’s (2020) results, where 69% of participants responded ‘Yes’ in the Speaker Unaware condition compared to 43% in the Speaker Aware condition [$\chi^2(1) = 14.9, p < .001$].

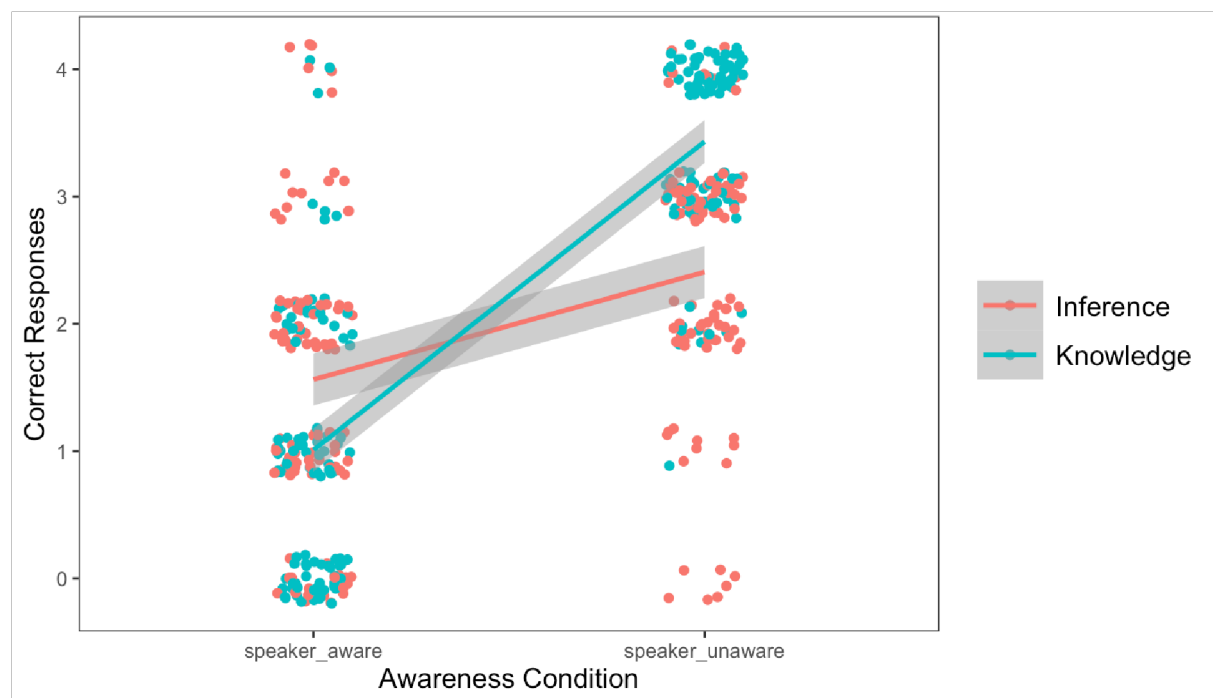
In both groups, participants were assigned to a counterbalanced list of speaker-aware and speaker-unaware versions of the stimuli. As participants read the stories in a different order (due to the counterbalancing), it was necessary to examine whether trial order explained variation on the Pragmatic Inference Task. Therefore, a full model with an interaction between the fixed effect of awareness and trial order was examined to measure whether this improved the model fit over a null model including the fixed effect only. In both the Knowledge [$\chi^2(1) = 1.77, p = .18$] and Inference [$\chi^2(1) = 2.36, p = .12$] Groups, the

model, including trial order, did not improve the fit, suggesting that the order in which the items were presented did not influence the results.

Finally, both groups were analysed together. A full model including the interaction between awareness and group improved the model fit significantly over a null model that omitted the interaction [$\chi^2(1) = 70.51, p < .001$]. Specifically, the effect of the awareness condition was larger for the Knowledge Group than the Inference Group [$\beta = 2.25, SE = 0.26, p < .001$]. Again, Trott and Bergen's (2020) results were replicated, although a slightly larger effect size was found [$\chi^2(1) = 28.1, p < .001$], which may be explained by the fact that this study had a larger sample size.

Figure 4. 7

Interaction Between the Awareness Condition and the Knowledge and Inference Groups



Overall, these results indicate that manipulating knowledge states (awareness) systematically changes individuals' ability to sample and maintain speaker information and deploy this information for pragmatic inference. The interaction between awareness and group (see Figure 4.7) reveals that the effect of altering speaker knowledge states is greater in the Knowledge versus the Inference conditions. These results indicate that individuals adopt a speaker's perspective to interpret the meaning of an indirect request.

Does mentalising predict responses in either the Knowledge or Inference Groups using different mentalising variables and added control variables? Partial replication and extension

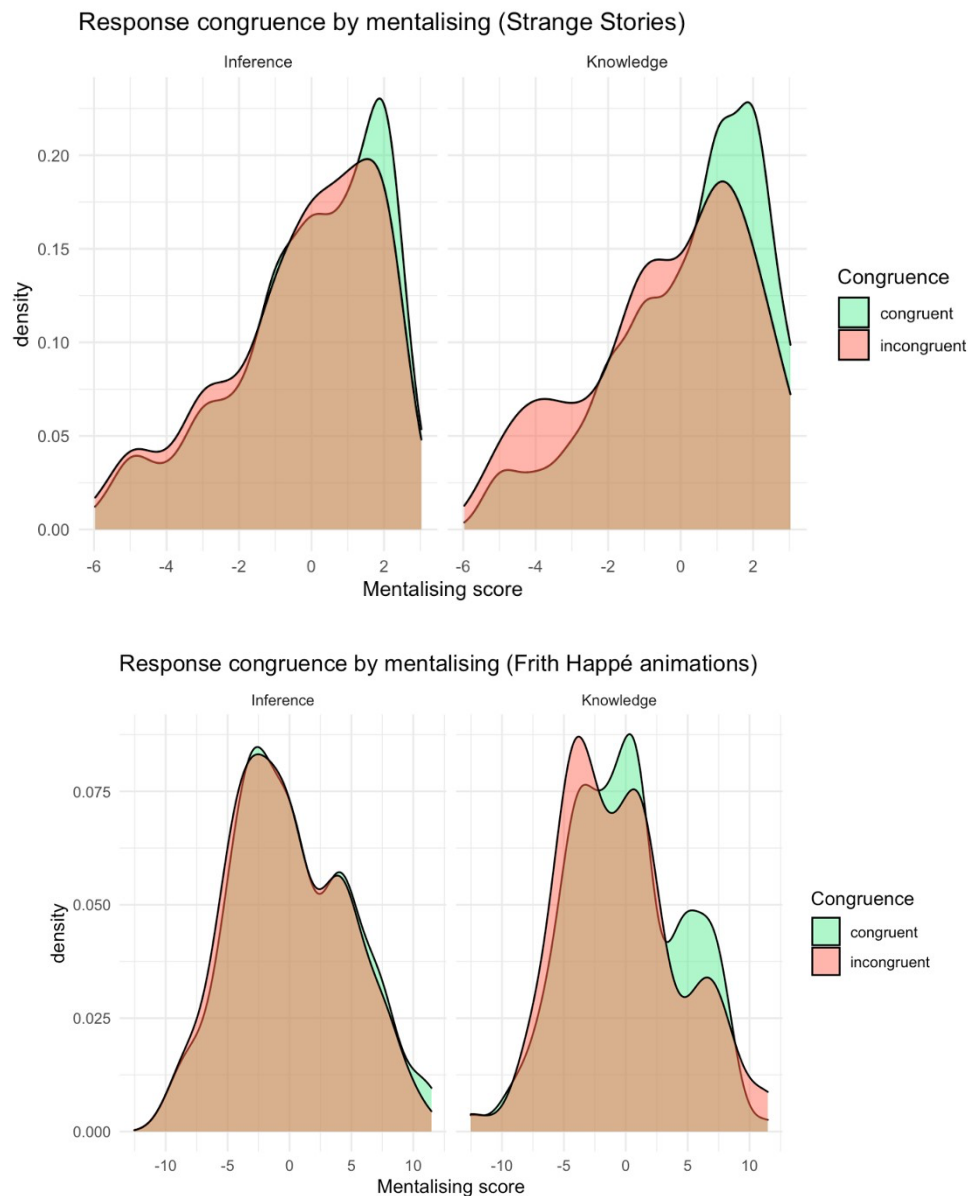
In the partial replication, the data were analysed to determine whether mentalising predicted performance above and beyond the control variables in the Knowledge or Inference Groups. In line with Trott and Bergen's (2020) results, this study predicted that better mentalisers will be more likely to choose a correct response in the Knowledge Group, after controlling for age, working memory and reading comprehension. However, mentalising ability was not anticipated to predict performance in the Inference Group. To compare the variables (both mentalising variables and controls), grand mean centring (i.e., taking each variable's scores and subtracting the mean of all the scores for that variable) was performed (Meteyard & Davies, 2020). Grand mean centring is used to improve interpretability and facilitate more meaningful comparisons between variables, and it also reduces multicollinearity. Without centring, the intercept in a linear model represents the predicted value when all predictors are at zero, which may not be meaningful (e.g., age). When the scores are grand mean centred, the intercept represents the predicted outcome at

the average value of the predictor, making the interpretation more meaningful. The results of the Strange Stories and the Frith-Happé Animations were analysed separately.

Figure 4.8 displays the distribution of mentalising scores across the Knowledge and Inference Groups for both the Strange Stories and the Frith-Happé Animations; a congruent answer in the Knowledge Group means correctly identifying that the speaker is aware of the obstacle in the aware condition and unaware condition. In the Inference Group, a congruent answer means identifying that it is more likely that the speaker is making a request in the unaware condition and that it is more likely that the speaker is not making a request in the aware condition.

Figure 4. 8

Density Charts of Individual Differences in Mentalising Ability Across Both Experimental Groups



Note: This density chart shows the distribution of mentalising scores across responses on the Pragmatic Inference Task in both groups. A density chart is a variation of a histogram; the peaks of the density plot indicate where the values are concentrated.

Mental State Reasoning (Strange Stories) and Pragmatic Inference Task

In the Knowledge Group, a full model, including the interaction between speaker awareness (awareness) and mental state reasoning (Strange Stories), explained significantly more variance [$\chi^2(1) = 7.46, p = .006$] than the null model without the interaction ($p = .414$). This can be interpreted as meaning that better mentalisers were more likely to correctly identify what the speaker knew or did not know (e.g., in the Speaker Unaware condition, better mentalisers were more likely to correctly identify that the speaker did not know the heater was broken).

As all the control variables correlated with the mental state reasoning variable ($r = .11 - .46$), a full model was created that included interactions between speaker awareness and the mentalising and control variables. This was then compared to a model that omitted the critical interaction between the speaker awareness condition and mentalising. The results indicate that the full model was a better fit for the data [$\chi^2(1) = 4.5, p = .034$] than the null model ($p = .493$), indicating that better mentalisers were more likely to correctly answer ‘No’ (i.e., no, he is not aware of the obstacle) in the Speaker Unaware condition [$\beta = -.24, SE = 0.11, p = .030$].

In the Inference Group, a full model including the interaction between speaker awareness (awareness) and mental state reasoning (Strange Stories) did not explain significantly more variance [$\chi^2(1) = 2.24, p = .135$] than the null model without the interaction ($p = .069$). Overall, these results suggest that the text-based measure of mental state reasoning predicts individuals’ ability to sample and maintain speaker knowledge states but not their ability to deploy this information for pragmatic inference. This relationship is maintained when controlling for reading comprehension, working memory, and age.

When using the Strange Stories as a mentalising predictor, the results of Trott and Bergen (2020) were replicated, i.e., mentalising predicts performance in the Knowledge Group but not the Inference Group.

Attributing Intentions to Shapes Task (Frith-Happé animations) and Pragmatic Inference Task

In this analysis, a full model including the interaction between speaker awareness (awareness) and mental state reasoning (Frith-Happé Animations Test) did not explain significantly more variance than the null model without the interaction in either the Knowledge Group [$\chi^2(1) = 2.24, p = .135$] or the Inference Group [$\chi^2(1) = 0.28, p = .594$]. Therefore, these results suggest that individuals who are better at attributing intentions to shapes are not necessarily better at sampling and maintaining speaker knowledge states or at interpreting the meaning of an ambiguous, indirect request more appropriately. In the Knowledge Group, there were significant main effects of reading comprehension [$\beta = .14, SE = 0.06, p = .019$], but the interaction between reading comprehension and the awareness condition did not reach significance. These results indicate that reading comprehension skills may have a greater role in sampling and maintenance than pragmatic inference. No further analysis was conducted.

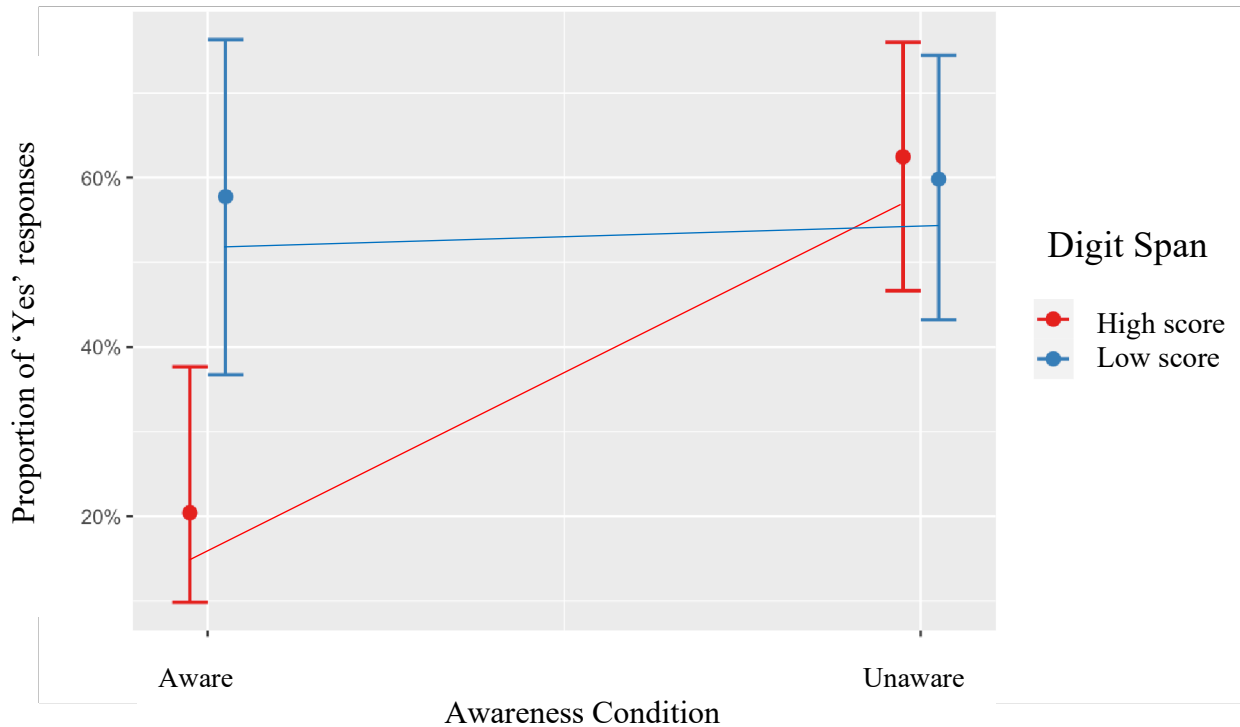
These results do not replicate those of Trott and Bergen (2020) and suggest that mental state reasoning (Strange Stories) and attributing intentions to shapes (Animations task) involve different components of mentalising.

Post-Hoc Analysis

To examine whether any of the mentalising predictors or the control variables explained variance in the Inference Group, a mixed model was created with both mentalising variables (Strange Stories and the Frith Happé Animations) and all three control measures (working memory, reading comprehension, age). The only significant interaction was found between working memory and speaker awareness [$\beta = -.25$, $SE = 0.09$, $p = .006$]. In the Inference Group, a full model including the interaction between speaker awareness (awareness) and working memory (Digit Span) explained significantly more variance [$\chi^2(1) = 7.44$, $p = .006$] than the null model without the interaction ($p = .656$); this suggests that variance in the ability to deploy speaker information to interpret an indirect request is supported by working memory (see Figure 4.9). More specifically, individuals with better working memory skills were more likely to select the appropriate response (i.e., ‘No’ he is not making a request) when the speaker was aware of the obstacle to fulfilling the request. Still, they were as likely as individuals with poorer working memory to select the appropriate response when the speaker was unaware of the obstacle.

Figure 4. 9

Interaction Between Awareness and Working Memory in the Inference Group



To examine whether any of the mentalising predictors or the control variables explained variance in the Knowledge Group, a mixed model was created with both mentalising variables (Strange Stories and the Frith Happé Animations) all three control measures (working memory, reading comprehension, age). Other than the significant interactions found between the Strange Stories and speaker awareness, the only significant interaction was found between reading comprehension and speaker awareness [$\beta = -.27$, $SE = 0.01$, $p = .014$]. In the Knowledge Group, a full model including the interaction between speaker awareness (awareness) and reading comprehension (Strange Stories Physical) explained more variance [$\chi^2(1) = 5.83$, $p = .015$] than the null model without the interaction ($p = .101$). This suggests that participants with better reading comprehension abilities were better at sampling and monitoring information about whether the protagonist knew about an obstacle to fulfilling a request or not.

4.4 Discussion

Effects of manipulating speaker knowledge on indirect request comprehension (pure replication)

This study examined how speakers' knowledge states affect a comprehender's interpretation of an indirect request in the context of a written narrative, and whether mentalising is involved in this process. The first stage of the study was a pure replication of Trott and Bergen's (2020) Experiment 1. The results showed that speaker knowledge modulated all stages of pragmatic inference (sampling, maintenance, and deployment), but participants sampled and maintained information about speakers' knowledge states at a higher rate than they deployed it for pragmatic inference. As the results show, the speaker's knowledge states affect sampling, maintenance, and deployment differently, supporting Apperly's (2018) component-based account of perspective-taking. These results indicate that when reading narratives, individuals adopt the speaker's perspective and use information about what the speaker knows or believes to interpret ambiguous requests. These findings offer new avenues for investigating the cognitive mechanisms involved in interpreting pragmatic inference in text. For example, it is possible that mentalising may support taking a character's perspective differently during each stage (sampling, maintained, and deployment).

Does mentalising predict responses in either the Knowledge or Inference Groups using different mentalising variables and added controls (partial replication and extension)?

The second stage of the study explored whether mentalising supports sampling and maintenance, or deployment of characters' perspectives. Trott and Bergen (2020) used a text-based mentalising measure, the Short Story Task, to demonstrate that mentalising supports building a mental model of the speaker's perspective during sampling and maintenance, but the ability to deploy this information to interpret the meaning of an indirect request relies on other skills. The current study was a partial replication, as different mentalising measures (Strange Stories and Frith-Happé Animations) were used. Different mentalising measures were added as concerns have been raised about the Short Story Task's format, coding and psychometric properties. More specifically, its format is very similar to a traditional reading comprehension test, and there is no attempt within the coding to describe what type of mentalising concepts are being measured (Turner & Felisberti, 2017). We also included a mentalising task that does not draw on language comprehension, the Frith-Happé Animations, to reduce the confounding effects of verbal abilities given that both the Strange Stories and the Pragmatic Language Task are text-based. To extend Trott and Bergen's (2020) results, in addition to a reading comprehension control, extra controls for working memory (Digit Span) and age were added. Working memory was also included as a control measure because research indicates that people can coordinate communication by becoming sensitive to the perspectives of others (through attention cuing and memory retrieval), without the need to fully represent the speaker's perspective (Horton & Grieg, 2005; Long et al., 2018; Ryskin et al., 2015; Wardlow, 2013).

This study predicted that both the mentalising measures would predict performance when individuals sampled and maintained information from the text but not when they deployed it for pragmatic inference. This pattern of results was found using the text-based measure of mental state reasoning (Strange Stories) but not using the Attributing Intentions to Shapes Task (Frith-Happé Animations). In light of research that has revealed a lack of

coherence between mentalising measures (Bialecka-Pikul et al., 2021; Warnell and Redcay, 2019), it can be reasoned that intrinsic differences in the tasks' structure may be responsible for these differing result patterns. Per the taxonomy presented in Chapter 3, these tasks measure different aspects of mentalising (Frith-Happé Animations – attributing mental states to objects, Strange Stories – pragmatic language) and recruit different non-mentalising skills (visuospatial reasoning vs reading comprehension respectively). The Strange Stories test also requires the participant to interpret the mental states of people in social contexts and draw on cognitive and affective aspects of mentalising (Bosco et al., 2016). The Short Story Task draws upon the same skills, which possibly explains why both tasks predict performance on Trott and Bergen's Pragmatic Language Task. However, the results of Study 1 suggest that adults' performance on the Strange Stories and Short Story tasks are uncorrelated, implying either that they assess different mentalising abilities or psychometric weaknesses. It is possible, then, that the association of both Strange Stories and Short Stories with the pragmatic language task is explained by shared verbal demands.

Strange Stories also draws on a range of pragmatic language skills, which may explain the association between mentalising and the Pragmatic Language Task in this study. Six out of eight of the Strange Stories items ask why a character says something that is not literally true. Therefore, the participant must use contextual cues, such as the events in the text and the speaker's intentions to infer the meaning of the characters' statements. Although Strange Stories may draw upon pragmatic abilities, it also examines individuals' ability to interpret the mental states of characters in social situations (Bosco et al., 2017). More specifically, the task draws on mentalising abilities such as false beliefs and lies as well as attributing desires, beliefs, and intentions (Happé et al., 1998; White et al., 2009). Thus, it is uncertain whether the pragmatic or mentalising aspects of the Strange Stories (or synergy of both) are associated with performance on the Pragmatic Language Task (Bosco et al., 2017).

To investigate whether the confounding effects of language are responsible for the relationship between the text-based mentalising task (the Short Story task) and the Pragmatic Language Task in Trott and Bergen's study, this study also included the Frith-Happé Animations to measure mentalising in the absence of pragmatic language demands.

In contrast to Strange Stories, performance on the Animations task did not predict scores on the Pragmatic Language Task during sampling, maintenance, or deployment of pragmatic inference. There were stronger correlations between the mentalising measures and their non-mentalising control conditions (i.e., Strange Stories Physical and Frith-Happé Goal-Directed animations) than between the mentalising measures themselves. As such, the non-mentalising aspects of the Strange Stories (vocabulary, grammar, inferencing) and the Animations task (visual processing, expressive language) may explain task performance more than any shared mentalising aspects of the tasks.

The Animations task requires participants to attribute human intentions to the movement of non-human objects and then describe these interactions using mental state terms (intentionality score; Castelli et al., 2000; Wick et al., 2016). Researchers have noted large discrepancies between performance in narrative tasks (e.g., Strange Stories), which draw upon social understanding and language use in social settings, and tasks that attribute human emotions to objects (e.g., The Frith-Happé Animations), which mainly draw upon visual perception (Rasmussen & Jiang, 2019; Scholl & Goa, 2013). Although the Animations task elicits the perception of animacy in moving objects, research suggests that individual differences in task performance may be due to the verbal responses given, which would indicate that verbal abilities are also drawn upon in this task (Abel et al., 2000; New et al., 2010; White et al., 2011). Clinical groups that have shown poor performance in the Frith-Happé Animations perform as well as typically developing controls in animacy perception tasks that do not require verbal responses. New et al. (2010) revealed that autistic

children's performance was comparable to typically developing controls in a study that required participants to detect changes in blurred still photographs of social and non-social scenes. Rasmussen and Jiang (2019) posit that autistic individuals may be as capable as typically developing individuals of attributing intentions to shapes but have difficulties verbalising these intentions. Together, these studies suggest that the mentalising skills tested in the Strange Stories (mentalising concepts, pragmatic language, interpreting mental states in social situations) are different from those in the Frith-Happé Animations (attributing mental states to moving objects). Additionally, as the success in the Animations task relies on the individuals' expressive language and use of mental state vocabulary, it may not be a suitable measure to use in combination with the Strange Stories to reduce the confounding effects of verbal abilities.

Taken together, these results suggest that the mentalising concepts drawn upon by the Strange Stories predict sampling and maintenance of speakers' perspectives during pragmatic language tasks. However, further research is needed, as this association may also be due to the pragmatic language elements of the task. Trott and Bergen (2020) concluded that mentalising predicted sampling and maintenance using the Short Story Task, possibly indicating that mentalising concepts tapped by both tasks (intentional language and interpreting mental states in social situations) explain this association. However, we cannot be certain of this because the results of Study 1 hint that the Short Story Task is not a reliable mentalising measure due to its reliance on reading comprehension and associated abilities such as grammar, vocabulary, and inferencing (Cain et al., 2001). Thus, participants may fail this task due to poor comprehension of the text, even if they have the capacity to understand the mental state concepts within the task.

To further research the role of mentalising in pragmatic inference during reading, it would be useful to pair the Strange Stories with a task that controls for the confounding

effect of pragmatic language more effectively. The Silent Films Task (Devine & Hughes, 2013) uses silent films as stimuli, thereby avoiding any confounding effects of pragmatic language while assessing similar components of mentalising as Strange Stories. When examining the psychometric properties of the Strange Stories and the Silent Films Task, Devine and Hughes (2016) demonstrated that results on both tasks were strongly correlated even when verbal abilities and narrative comprehension were taken into account. Therefore, Silent Films may be a useful comparison task of mentalising to include in future replications of this study.

Deployment of pragmatic inference

Trott and Bergen's (2020) suggest, on the basis of the results of Experiment 1, that mentalising supports sampling and maintenance but that the deployment of pragmatic inference relies on other skills. The post-hoc analysis performed in the current study implies that working memory may have a key role in the deployment of pragmatic inference. The analysis identified an interaction between working memory and speaker awareness (whether they were aware or unaware of an obstacle to fulfilling a request) in the Inference Group, during deployment. Further analysis highlighted that individuals with better working memory skills were more likely to select the appropriate response when the speaker was aware of the obstacle to fulfilling the request. At the same time, they were as likely as individuals with poorer working memory to select the appropriate response when the speaker was unaware of the obstacle. This may indicate that, when interpreting literal comments in the aware condition, individuals with better working memory skills were better at sampling and maintaining the fact that the speaker had knowledge of the obstacle without being prompted to remember this information. As a result, they were able to use this

information more effectively to interpret the speaker's utterance (e.g., 'It's cold in here') as a literal comment about the state of affairs rather than as a request.

These results suggest that during the deployment of pragmatic inference, interpreting literal comments in the speaker-aware condition draws on working memory, whereas interpreting indirect requests in the unaware condition relies on other skills. Performance was similar in the unaware condition in individuals with low or high working memory scores, indicating that working memory is comparatively less important when interpreting indirect requests than when interpreting literal comments. It, therefore, remains unclear from the results of this study which cognitive skills support the interpretation of indirect requests. Interpreting indirect requests relies on contextual information. For example, if one reads the utterance 'It's cold in here' in isolation, the literal interpretation (a comment about the temperature) would most likely be taken (Gibbs, 1981). In contrast, when the speaker is unaware that the heater is broken as he gets into the car and says, 'It's cold in here', participants were significantly more likely to interpret this utterance as a request to turn on the car's heating. Therefore, it is probable that the reader uses information about the speaker's knowledge states to interpret the request function of the utterance.

This study did not find that mentalising predicted performance during the deployment of indirect requests. However, this may be because the methods used were not sensitive enough to elicit the effect of mentalising. Mentalising may work in combination with working memory, or other skills not measured in this experiment may allow a perceiver to understand the speaker's perspective, possibly rendering the effect of working memory and mentalising harder to uncover. Meta-analysis investigating pragmatic language in patients with traumatic brain injury has found that five key cognitive constructs (declarative memory, working memory, executive functions, attention, and mentalising) are significantly correlated with comprehension of pragmatic language (Rowley et al., 2017). Of these

constructs, declarative memory has the strongest correlation with pragmatic language. The perceiver may draw upon an aspect of declarative memory—namely semantic knowledge (e.g., cars have heaters)—during the deployment of pragmatic inference (Grodner & Sedivy, 2011). However, functional magnetic resonance imaging (fMRI) studies suggest that the mentalising network in the brain is still activated when interpreting indirect requests, even when contextual cues are available. Van Ackeren et al. (2012) presented participants with pictures (e.g., a man standing next to a window) while they listened to utterances (e.g., ‘It’s hot in here’), which activated cortical motor areas of the brain responsible for perceiving and planning action as well as areas associated with the mentalising network (the medial prefrontal cortex, praecuneus, and the bilateral temporoparietal junction). Their findings suggest that mentalising is needed to interpret the meaning of indirect requests in context.

Trott and Bergen (2020) also propose that the characteristics of the Pragmatic Language Task may mask the effects of mentalising during deployment. In the Knowledge Group, participants only have to sample and maintain information to answer the test items (e.g., does the speaker know the heater is broken?). In contrast, in the Inference Group, the participants must sample and maintain the speaker’s knowledge states and then deploy this information to make a pragmatic inference about the meaning of the utterance (e.g., is he making a request?). Therefore, the effects of mentalising may not have been elicited as some participants may have failed to sample and maintain information about the speaker’s perspective.

Conclusions

The first stage of the study examined whether altering the speaker’s knowledge states modulated individual’s interpretation of a pragmatic inference during each of the three

components (sampling, maintenance, deployment) of Apperly's component model of perspective-taking. Individuals sampled and maintained information about speakers' knowledge states at a higher rate than they deployed it for pragmatic inference. As each stage was affected differently by altering the speaker's knowledge, this indicates that these are different component processes. The pure replication results replicated Trott and Bergen's (2020) Experiment 1's results. The next stage of our experiment examined the way in which mentalising supported each component.

This study indicates that mentalising may be involved when sampling information about the speaker's perspective and maintaining this information in memory, but not when deploying this information to make a pragmatic inference. However, given the lack of clarity on what mentalising tasks measure (Chapter 3), it is still unclear whether this finding isolates the role of mentalising in pragmatic inference or whether the pragmatic language demands of the Strange Stories task explain the observed relationship. The results of this study suggest that the Frith-Happé animations do not capture aspects of mentalising recruited during pragmatic inference generation. A visual task that measures the same mentalising concepts as Strange Stories, such as the Silent Films Task, would provide a clearer control for the language demands of Strange Stories. This study also suggests that mentalising does not support deployment of the speaker's perspectives when making a pragmatic inference during reading. However, this may be due to the characteristics of the Pragmatic Language Task used here. To investigate the effect of mentalising on the deployment of pragmatic inference, it is necessary to use a more sensitive paradigm.

5. Does Mentalising Predict Deployment of Pragmatic Inference when Interpreting Ironic Comments?

5.1 Chapter overview

The focus of both Study 3 (see Chapter 4) and Study 4 was to investigate whether mentalising has a special role in processing pragmatic language embedded in written text. In fictional narrative texts, dialogue-based interactions between characters are often key to narrative development. The association between adult reading comprehension and mentalising reported in Study 1 (Chapter 2) may, therefore, be driven by individual differences in pragmatic language processing (see Figure 4. 1, Chapter 4, p.165). Study 3 illustrated that the ability to identify what a character in a text knows or thinks about the world and to hold this information in memory (i.e. the sampling and maintenance stages of Apperly's (2018) component model of perspective-taking) is supported by mentalising. The final stage of this process is deployment, i.e., the ability to use mental state information in combination with linguistic information to make a pragmatic inference. The results of Study 3 indicate that mentalising did not support the deployment of mental state information when

readers were deciphering the meaning of indirect requests. Notably, post-hoc tests revealed that participants with higher working memory scores were more able to accurately judge the intended meaning behind a speaker's utterances than those with lower working memory scores. Working memory may be particularly important when integrating mental state input, contextual information, and linguistic information from the utterance to infer the speaker's intentions. However, it is possible that the methods used in Study 3 were not sensitive enough to elicit the effects of mentalising in deployment for pragmatic inference. Notably, a number of other studies using various paradigms have found that mentalising supports deployment (Dumontheil et al., 2010; Keysar et al., 2003; Marocchini & Domaneschi, 2022; Van Ackeren et al., 2016). Therefore, Study 4 used different methods to explore whether mentalising has a role in the deployment of mental state information to interpret pragmatic language during reading.

5.1.1 Introduction

Apperly (2018) indicates that perspective sensitivity (e.g., monitoring perspectives through attention and memory) is not always enough to ensure comprehension during the deployment phase, and perspectives may need to be represented, which would recruit mentalising. He refers to his own work using a visual-based referential communication paradigm (the Director's Task) that attempts to separate sampling, maintenance, and deployment. Participants make a high number of errors deploying representations of visuospatial perspective when the director has a different perspective from their own (Dumontheil et al., 2010; Keysar et al., 2000). This is because when two people have two different visuospatial perspectives (e.g., Two people sitting opposite each other at a table see

the same cup of coffee but from different angles) individuals are biased towards their own perspective (Keysar, 2000). The evidence on the role of mentalising at the deployment stage is therefore mixed, it is possible that the experimental paradigm used in Study 3 was insufficiently sensitive to detect this relationship.

In addition, there is evidence from neuroimaging studies that the mentalising network (mPFC) is activated during the deployment of pragmatic inference. A study by Van Ackeren et al. (2016) used fMRI to determine whether the mentalising network is recruited during the deployment of highly indirect requests compared to two control measures. To do this, participants read dialogues between two people in which the context (speaker A's questions) was manipulated to systematically alter the meaning of speaker B's reply (which was always the same; see Table 5.1). Altering the context resulted in three experimental conditions: speaker B's response could be a factual 'direct reply' to speaker A's question, an 'indirect reply', or an indirect request for action. After hearing the dialogue between the two speakers, the participants had to indicate whether they thought that speaker B's response was a request for speaker A to act or not.

Table 5. 1

Example Dialogues from Van Ackeren et al. 's (2016) Stimulus Set

	Speaker A	Speaker B
Direct Reply	How far away is China?	It's quite far away.
<i>Coded meaning</i>		
Indirect Reply	Have you started preparing for the exam?	It's quite far away.
<i>Speaker meaning</i>		
Indirect Request	Shall I move the TV closer to the sofa?	It's quite far away.
<i>Speaker meaning + action</i>		

Note: A given reply ('It's quite far away') is interpreted as a direct reply, an indirect reply, or an indirect request, depending on the question.

The results of this study indicate that making a pragmatic inference about the meaning of an indirect request engages areas (mPFC) associated with the mentalising network, providing further evidence that mentalising is involved in deployment. However, the fMRI results do not indicate which aspects of mentalising may be involved or which associated skills (e.g., reading, or pragmatic skills) may also support indirect request interpretation in this task (Schaafsma et al., 2015; Warnell & Redcay, 2019). Therefore, although Van Ackeren et al.'s (2016) results do suggest that mentalising is involved in deployment, they do not allow specification of the types of mentalising that support deployment.

In Experiment 2 of Trott and Bergen's (2020) study, they suggest that mentalising is recruited when people are explicitly directed to attend to the speaker's mental states. In their Experiment 1 (replicated here in Study 3, Chapter 4), the influence of mentalising was detected in the Knowledge Group—targeting sampling and maintenance—because people were explicitly asked to attend to the mental states of characters (e.g., 'Does he know the

heater's broken?'). In contrast, in the Inference Group—targeting deployment— participants were only implicitly asked to attend to mental states (e.g., 'Is he making a request?'). In Experiment 2 of Trott and Bergen's (2020) study, the key pieces of information in the narrative regarding the speaker's perspective were made more explicit by adding descriptive text to render the intended meaning more salient (e.g., 'It's cold in here; **too bad the heater is broken**') and placing the text in bold to ensure that the reader sampled and maintained this information. After making these changes to their paradigm, they found that mentalising indeed supported deployment of the speaker's mental state information during pragmatic inference.

Trott and Bergen's proposal that mentalising is only recruited when readers are explicitly directed to attend to mental states is partly supported by Back and Apperly (2010), who found that participants only monitored characters' false beliefs when explicitly directed to do so. However, Apperly (2018) hypothesises that there are other factors that may cause comprehenders to make mental state inferences spontaneously, such as motivation and individual differences in cognitive resources (e.g., mentalising, working memory; Surtees et al., 2016).

The current study aims to further investigate whether mentalising is involved in the deployment of pragmatic inference. Study 3 focused on one form of social pragmatic language: highly indirect requests. The current study investigated the role of mentalising in the deployment stage in comprehension of another type of pragmatic language, namely ironic comments (Marocchini & Domaneschi, 2022).

5.1.2 Ironic comments

Pragmatics examines the intentions behind a verbal utterance (spoken or in text) within a dialogue between two people. The intentions behind a pragmatic utterance may be to amuse, inform, or direct (Pershey, 1980). Irony is a form of pragmatic language that is commonly understood as an utterance with a literal meaning that is the opposite of the utterance's intended meaning. Often, irony is used to achieve a dryly humorous effect. However, there is no standard definition of irony in psychological research, and researchers suggest there may be various subtypes of irony (Gibbs & Colston, 2007). In this study, irony is defined as an utterance (1) that has a literal meaning that is incongruent with the intended meaning, (2) that is intended to amuse, (3) that has implicit meaning, and (4) that needs to be evaluated to distinguish between an ironic and non-ironic (literal) interpretation (Burgers et al., 2011). Although irony is often thought of as dryly humorous, the function of ironic comments can include gentle teasing, which has been shown to occur between parents and children (Dunn & Brown, 1994), friends (Keltner et al., 1998), and romantic partners (Baxter, 1992).

Interpreting irony is a complex process, as the listener needs to understand that the literal meaning of a speaker's utterance is incongruous with the intended meaning, that this incongruity is deliberate, and that the intention of the speaker is not to mislead the listener but to entertain them (Hoicka et al., 2008). One of the key psycholinguistic theories used to explain how this process works is the incongruity-resolution theory (Suls, 1977). This account proposes that an ironic comment is interpreted in two stages. The listener must first detect an incongruity between the literal meaning of an utterance and the speaker's intended meaning and then must resolve the incongruity by retrieving information from their own background knowledge (Suls, 1983). However, this theory does not account for the fact that

an incongruent utterance is not always intended to be ironic because individuals can lie, speak mistakenly, or hold false beliefs. A key aspect of ironic comments not addressed by incongruity-resolution theory is that there may be individual differences in the listener's ability to infer the speaker's true intentions when interpreting ironic comments. These individual differences may be related to their mentalising abilities.

Theories of humour have suggested that mentalising may be necessary to interpret the intention of the speaker as ironic. Building on the incongruity-resolution theory, the Mind Reading Hypothesis (Howe, 2003) suggests that it is this resolution of the incongruity between what was said and what was meant that is the source of amusement. The enjoyment of the ironic comment comes at the point when the speaker perceives that the listener has resolved the incongruity, and the listener perceives the intentions behind the speaker's utterance. At this point, both the speaker and listener are aware of the cognitive and emotional response intended by the ironic comment.

In spoken dialogue, communicators monitor their conversational partner's perspectives to frame their responses appropriately, but it may not be necessary to fully represent a speaker's perspective to interpret them correctly (Horton & Gerring, 2005; Sperber & Wilson, 2022). More often, it may be the case that communicators can coordinate conversations and become sensitive to each other's perspectives through attention and memory (Horton & Gerring, 2005). For instance, if two siblings are watching a boxing match and one says ironically, 'Dad would enjoy this', the listener may recognise that they do not mean this literally as the listener may remember that their father hates boxing, or they might notice that their sibling is smiling while making this comment. In this case, we might expect the deployment of an ironic utterance to be supported by accessing shared knowledge through memory. However, if the listener had a different perspective from their sibling (e.g., they believe their father likes boxing), then shared knowledge and attention to facial cues

may not be enough to interpret the comment correctly. At this point, the mental states of the conversational partner must be fully represented to understand what the intended meaning of the speaker's utterance is and then to resolve any misunderstandings if necessary (Apperly, 2018). In this case, mental state reasoning may be recruited during deployment.

In research, the processing of ironic comments has been studied more commonly in spoken language than in text, and it is possible that readers process ironic comments in text differently than in real-life spoken interactions (Bryant & Fox Tree, 2002; Gibbs, 2000). Reading removes many of the cues that we use to understand the perspectives of others, such as facial expressions, eye gaze, laughter, and tone of voice (Apperly, 2018). Moreover, there is no way of 'repairing' an ironic utterance if it is not understood in the text other than re-reading (Burgers et al., 2012). To understand the intended meaning of an utterance, we must rely on our understanding of the surrounding written context. This may explain why the results of Study 3 were that mentalising supported sampling and maintenance of mental state information but not deployment. Representing the character's thoughts and feelings may be more important at this stage when reading spoken dialogue. Study 3 focused on highly indirect requests; however, irony in written texts may be processed in a similar way. For example, if we read the utterance 'last night was a blast' alone without any surrounding context, we have no way of knowing whether this comment was intended to be taken literally or ironically. We need to know the speaker's attitudes towards the previous night to interpret whether this statement is incongruous with their true feelings, and we would gain this information from the surrounding written context. When deploying this information to make a pragmatic inference, it may simply be necessary to remember the speaker's attitudes rather than fully represent them to detect the presence of irony, which would fit with the finding that working memory predicts deployment reported in Study 3.

Although highly indirect requests and ironic utterances are both non-literal statements, the key difference between them is that when making an indirect request, the speaker is not deliberately uncooperative. In contrast, when speaking ironically, the speaker is intending to mislead the listener for comic effect. This may mean that ironic utterances are more cognitively taxing and may rely more on mentalising than indirect requests. This idea is supported by evidence from a recent Italian study showing that mentalising, as measured by the Strange Stories, supports the deployment of ironic utterances (Marocchini & Domaneschi, 2022; see Section 5.1.4).

5.1.3 Linguistic vs social pragmatic language

Recent research has indicated that pragmatic interpretive strategies are required depending on the specific features of an utterance and individual differences in mentalising ability. Andres-Roqueta and Katos (2017) suggest that pragmatic language takes two forms: social and linguistic. Comprehension of linguistic pragmatic utterances relies on the listener's knowledge of pragmatic norms and linguistic cues; an example of a linguistic pragmatic form is a conventionalised indirect request. Conventionalised indirect requests are utterances that do not have a literal directive meaning but have been accepted within language as such. For example, suppose each word in the following utterance, 'Can you pass the salt?', is taken literally. In that case, the utterance is understood as a non-directive, literal question that asks if you have the ability to perform the act. However, the use of 'Can you' at the start of a sentence has been accepted in English as a polite way of requesting action (i.e., rather than just saying 'Pass the salt'). In fact, research shows that due to the common use of 'Can you...' at the start of request sentences, the conventionalised indirect request

interpretation is more likely to be taken than the literal meaning of the sentence (Gibbs, 1983). So, when a listener hears the linguistic signal ‘Can you’ at the start of a sentence, they more often expect a request for action to be forthcoming. Linguistic indirect requests are less dependent on understanding the speaker's perspectives and the surrounding context and are not thought to recruit mentalising (Feng et al., 2021; Andrés-Roqueta & Katsos, 2020).

Contrastingly, social pragmatic utterances, such as highly indirect requests or ironic comments, are dependent on understanding the speaker’s perspectives and the surrounding context and are thought to be supported by mentalising. For instance, imagine someone at a dinner party dropped the salt grinder and shattered it; a speaker might say, ‘Can you pass the salt?’ to lighten the mood. In this case, the speaker intends the utterance to be ironic. Social pragmatic utterances do not contain linguistic cues, and therefore the full utterance needs to be heard in context to gain understanding (Reboul, 2004). Empirical evidence supporting the distinction between task-specific mentalising-based social pragmatics and non-mentalising based linguistic pragmatics is reported in a correlational study by Andres-Roqueta and Katos (2020). This study examined the association between mentalising (measured with a higher order false belief task) and structural language with two pragmatic language tasks: one that primarily drew on linguistic abilities (linguistic pragmatic task) and another that drew on linguistic abilities and mentalising (social pragmatic task) in autistic children, children with developmental language disorder, and typically developing controls. Their results indicated that scores on the linguistic pragmatic task were predicted by structural language in all three groups. However, scores on the social pragmatic task were predicted by both structural language and mentalising. Although more research is needed, these findings indicate that different skills support linguistic and social pragmatic skills.

5.1.4 Investigating the differences between social and linguistic pragmatics

Following Andres-Roqueta and Katos' (2017, 2020) work, Marocchini and Domaneschi (2021) investigated whether the distinction between linguistic and social pragmatics can be observed in typically developing adults. To do this, they examined the association between mentalising and a form of linguistic pragmatic utterance (conventionalised indirect requests) in comparison with a form of social pragmatic utterance (ironic comments). Ironic comments were used given that previous research has proposed that their interpretation recruits mentalising (Deliens et al., 2018). Non-directive literal comments were used as a comparative control, because they are not thought to recruit mentalising. In a self-paced, region-by-region reading paradigm, participants read a set of narrative items, comprising three conditions. In each narrative item (see Table 5.2), the initial context-setting section and the final target sentence remained the same, but the second part of the context was manipulated to change the interpretation of the target utterance, creating three conditions (conventionalised indirect requests, ironic comments, and non-directive literal questions).

Table 5. 2

Example Narrative Item and the Three Conditions in Marocchini & Domaneschi's (2021) Study

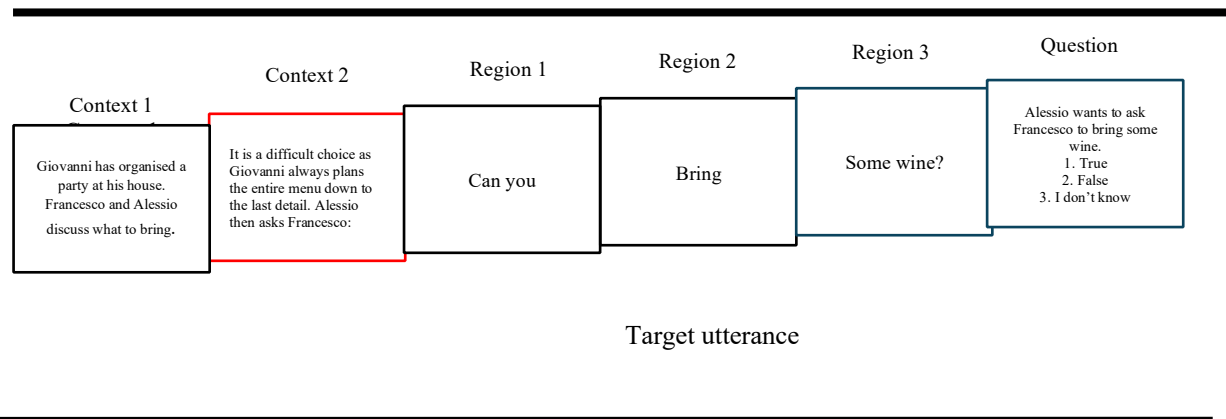
Condition	Context		Target Sentence
	Part 1	Part 2	
Indirect Request	Giovanni has organised a party at his house. Francesco and Alessio discuss what to bring to the party.	It's a difficult choice because Giovanni always plans the entire menu down to the last detail. Alessio asks Francesco:	Can you bring the wine?
Literal Question		It's a difficult choice because the trunk of Francesco's motorbike is almost full. Alessio then asks Francesco:	
Ironical comment		It's a difficult choice, because Giovanni does not drink any alcohol, and he is frequently teased by his friend because of it. Alessio jokingly says to Francesco:	

Note: Example of a narrative item with the three conditions (conventionalised indirect request, nondirective literal question, ironic comment). In each condition, Part 1 of the context and the target sentence stay the same, whereas Part 2 of the context is the manipulated variable.

The narrative items were presented across five screens: the context (Part 1 and Part 2) and the target utterance across three screens (see Figure 5.1). Participants read the context at their own pace, and their reading times were recorded across the three regions of the target sentence.

Figure 5. 1

Example Screenshots of the Indirect Request Condition in Marocchini and Domaneschi's (2022) Study



Note: In this paradigm, only the context 2 screen was manipulated.

The target sentence was split into three regions to identify whether the individuals comprehended the intended meaning of the sentence (e.g., indirect request, literal comment, or ironic comment) at a local level before the whole sentence was read or at a global level once the whole sentence was read. Longer reading times point to increased processing and indicate when comprehension of the intended meaning has been achieved. For example, if longer reading times were found in region 1 (e.g., 'Can you') and shorter reading times in regions 2 and 3, this would suggest that the individual has identified the intended meaning of the sentence (e.g., that it is a literal question) at the start of the utterance. However, if regions 1 and 2 were read at a similar pace but region 3 was read more slowly, this could suggest that the intended meaning of the sentence was comprehended at the end of the sentence after integrating the meaning of the words across each region of the target utterance. Longer reading times at the end of the sentence indicate 'wrap-up effects' in

integrating information to attain the intended meaning (Rayner et al., 2000). Pause length in wrap-up effects may be determined by the information density of the sentence and/or the reader's processing capacity (Haberlandt & Graesser, 1985; Stine-Morrow et al., 2010).

Although research has indicated that irony is more cognitively taxing than the other types of utterance (Bryant, 2012), Marocchini and Domaneschi's (2022) attempted to control for the information density of target sentences in each condition. A norming study was conducted to attain parity in the target sentences' length, frequency, and readability. Thus, remaining differences in reading times should reflect individual differences in processing capacity. In their analysis, Marocchini and Domaneschi found that indirect requests, literal comments, and ironic comments were processed differently. Locally, region 1 (e.g., 'Can you') was read significantly more slowly in the indirect request and ironic comment condition than the literal question condition. At regions 2 ('bring') and 3 ('some wine'), the ironic comment condition was read significantly more slowly than the indirect request or literal question conditions. These results suggest that the intended meaning of the indirect requests was interpreted after the individuals read the tag 'Can you' at the start of the sentence, but the ironic comments needed to be read in full in order to derive the intended meaning. In addition, mentalising predicted faster reading times at region 1 of the indirect request condition ('Can you') but not region 2 or 3. This finding was unexpected and indicates that the better mentalisers were better at recognising linguistic cues that signal directive meaning. However, the association between the mentalising predictor (Strange Stories, a text-based mentalising task) and the linguistic cues could also be explained by the verbal elements of the tasks, which were not controlled for. In contrast, as expected, mentalising predicted faster reading times across each region of the utterance in the ironic condition as well as global reading times. These results indicate that mentalising is recruited to support

the meta-representation of the utterance both while the sentence is unfolding (across each region, locally) and across the whole sentence (globally).

Overall, these results support Andres-Roqueta and Katos' (2017) assertion that pragmatic interpretive strategies change depending on the specific features of an utterance and individual differences. Social pragmatic utterances are processed globally and are supported by language and mentalising, whereas linguistic pragmatic utterances rely on linguistic cues.

5.1.5 Mentalising measures and controls

The findings of Study 3 indicated that mentalising predicted sampling and maintenance of social indirect requests in the text-based Pragmatic Inference Task when using the Strange Stories, but not the Frith-Happé Animations, as an index of mentalising. As recent research presents mentalising as a multidimensional construct that interacts with other non-mentalising abilities (e.g., language, attention, memory), the reason for the different results may be due to the properties of the mentalising tasks that were used (Warnell & Redcay, 2019). Firstly, text-based Strange Stories may tap different components of mentalising from visuospatial Frith-Happé animations. According to the taxonomy (Chapter 3, Section 3.3.5), both tasks tap cognitive mentalising (inferences about thoughts and beliefs). However, the Frith-Happé animations measure the ability to attribute mental states to objects, whereas the Strange Stories measure the ability to attribute mental states to humans. Research suggests that it is possible to attribute human intentions to inanimate objects and humans (Johnson & Barrett, 2003). However, humans may have adapted to become particularly adept at recognising human intentions to work cooperatively with others

(e.g., protection from predators and food sharing), as well as individually learning to deceive and manipulate others to gain an evolutionary advantage (Brüne & Brüne-Cohrs, 2006; Byrne, 1996). Therefore, the Strange Stories may have revealed a relationship between recognising human intentions and pragmatic inference.

Alternatively, it may be the non-mentalising skills (e.g., verbal abilities) that explain the relationship between performance on the Strange Stories task and the Pragmatic Inference Task. Per the taxonomy, the Strange Stories draws on comprehension of various pragmatic concepts (e.g., persuasion, figure of speech) that could explain its association with the Pragmatic Inference Task. In Study 3, the Physical Stories task, which tests understanding of physical (rather than mental) states, was added as a control for reading comprehension, but not specifically pragmatic language comprehension. Therefore, in Study 4, to achieve a clearer picture of whether mentalising predicts social indirect request processing during reading, an additional mentalising task was included.

To compare mentalising tasks with and without verbal demands, a second non-verbal mentalising task that drew upon similar mentalising abilities was paired with the Strange Stories, namely the Silent Films Task. These tasks were paired together because the Silent Films Task was designed to draw on the same mentalising concepts tapped by the Strange Stories but without the verbal task demands (Devine & Hughes, 2013). By using the combination of the Strange Stories and the Silent Films Task, it is possible to characterise any associations found between the mentalising and pragmatic language comprehension during reading in more detail. For instance, if a relationship is found between the Strange Stories but not the Silent Films and the Pragmatic Inference Task, it may be that the associations are explained by shared language comprehension demands rather than mentalising.

5.1.6 Aim of the present study

Study 4 aims to investigate the role of mentalising in deployment of pragmatic inference during reading, by focusing on the way in which ironic comments are processed in comparison to two control conditions, conventionalised indirect requests and non-directive literal questions. This study builds upon Marocchini and Domaneschi's (2021) experimental design, which examined the involvement of mentalising in deployment of pragmatic inference in shorter narratives in the Italian language. The current study adapts this paradigm to examine the involvement of mentalising in the deployment of pragmatic inference in longer narratives in the English language.

To do this, the materials used in Study 3 from Trott and Bergen (2020) were adapted. As in Study 3, a self-paced reading paradigm was used. However, each narrative item was altered to create three conditions: ironic comments, conventionalised indirect requests, and literal questions (see Table 5.3). After reading the narratives, the participants were presented with the target utterance divided across three regions. Reading times were measured to characterise how each region of the target utterance is processed. Reading times were used (rather than a forced choice 'yes or no' question, as in Study 3) because Marocchini and Domaneschi (2021) indicate that measuring reading times may be more sensitive in characterising the ways in which different pragmatic forms are processed. Additionally, measuring reading times enables more precise identification of the involvement of mentalising in social pragmatic inferences (here ironic comments) in comparison to the two control conditions. To investigate the involvement of mentalising during the deployment of pragmatic inference in ironic comments and two control conditions, the Strange Stories and Silent Films Task were used. To control for shared verbal properties in the Strange Stories and the Pragmatic Inference Task, a measure of reading comprehension (ART-2) was

included. Additionally, as working memory was found to play a prominent role in deployment in Study 3, a forward and backwards digit span task was included to control working memory.

5.1.7 Research questions and predictions

RQ1. How are ironic comments (social pragmatic) processed in comparison to conventionalised indirect requests (linguistic pragmatic) and non-directive literal questions at the local and global levels?

- At the start of the sentence (region 1 – local level), it is predicted that no significant reading time differences will be found between ironic comments and conventionalised indirect requests. However, it is predicted that ironic comments will be read significantly more slowly than non-directive, literal questions.
- In the middle of the sentence (region 2 – local level), it is predicted that ironic comments will be read significantly more slowly than conventionalised indirect requests and non-directive literal questions. It is predicted that no significant reading time differences will be found between conventionalised indirect requests and nondirective literal questions.
- At the end of the sentence (region 3 – local level), it is predicted that ironic comments will be read significantly more slowly than conventionalised indirect requests and nondirective literal questions. Additionally, it is predicted that there will be no significant reading time differences between conventionalised indirect requests and non-directive literal questions.

- Across the whole target sentence (global level), it is predicted that ironic comments will be read significantly more slowly than conventionalised indirect requests and nondirective literal questions. It is also predicted that no significant reading time differences will be found between conventionalised indirect requests and nondirective literal questions.

RQ2. Does mentalising will predict individual differences in the ironic comment condition?

- Ironic comment condition: Mentalising will predict faster reading times across each part of the sentence (regions 1–3) and faster reading times across the whole sentence (global level).
- Linguistic comments and non-directive literal questions: Mentalising will not predict faster reading times across any part of the sentence (regions 1–3) and mentalising will not predict faster reading times across the whole sentence (global level).

5.2 Method

Participants

Participants were recruited online via two platforms: Prolific and SONA, the York St. John University online recruitment service. Prolific participants were paid £7.50 for participation, and SONA participants were given course credits. There were four sets of pre-screening criteria: (1) age (participants had to be over 18), (2) nationality (participants had to come from the United Kingdom or Ireland), (3) first language (participants had to speak English as a first language), and (4) no literacy difficulties.

The sample size was worked out on G-Power. The effect size ($f^2 = 0.12$) was computed using the data (e.g., sample size, number of groups and measures) from Marocchini and Domaneschi's (2022) paper and the 'Compute required effect size' function in G-Power). For two predictors and two control variables, it was estimated that for the planned analysis (power = 0.8, $\alpha = 0.05$, $f^2 = 0.12$), we would need the estimated number of participants to be between 90 and 100. In the end, data from were collected from 110 participants. Data from 20 participants were excluded from the study as they had an accuracy frequency of less than 75% on the Pragmatic Language Task. This left 90 participants in total. Ages ranged from 18 to 40 years old ($M = 26.5$, $SD = 6.56$). There were 56 females and 34 males.

Materials

The Pragmatic Language Task

In line with Marocchini and Domaneschi's (2022) study, the pragmatic form of interest (ironic comments) was examined in comparison with two control phenomena (conventionalised indirect requests and literal questions). Ironic comments are thought to require meta-representation and to recruit mentalising, whereas conventionalised indirect requests are processed through lexical-based strategies.

In this task, there were 15 narratives (12 target, 3 control) that describe daily life interactions between two or three characters, and each narrative ends with a target utterance. Each of the 12 target items had three experimental conditions (ironic comment, conventionalised indirect request, and literal question). In each item, the first part of the narrative (Context A) and the target utterance stayed the same. In contrast, the second part of

the narrative (Context B) was manipulated to produce the three different interpretations of the target utterance (see Table 5.3). Three counterbalanced lists were created so that each participant only saw one condition from each item; this was done to compare the processing times of ironic comments in comparison to the two control measures of conventionalised indirect requests and literal questions. The length of each narrative varied from 100 to 234 words in total (see Appendix E).

Table 5. 3

Example Item from Study 4’s Stimulus Set, Including All Three Conditions

CONTEXT A		
You and your friend are exhausted and sweaty from playing tennis. After playing, you invite him back to your apartment to hang out.		
CONTEXT B		
LITERAL QUESTION	IRONIC COMMENT	INDIRECT REQUEST
When you return to your apartment, you both see a notice taped to your apartment door: unfortunately, the water in your apartment building will be shut off for the next 24 hours. Your friend seems disconcerted by the news. You tell him not to worry, they probably haven't turned everything off just yet and you'll have to test it .	When you return to your apartment, you both see a notice taped to your apartment door: unfortunately, the water in your apartment building will be shut off for the next 24 hours. Your friend seems disconcerted by the news.	When you return to your apartment, you go to your room and privately check your email. You see that your landlord has sent you a message: unfortunately, the water in your apartment building will be shut off for the next 24 hours. Your friend, of course, doesn't know about this yet.
Once inside the kitchen, you head to the kitchen sink. Your friend turns to you inquiringly and asks,	Once inside the kitchen, your friend, still drenched in sweat, turns the knob on the kitchen tap, but the water does not flow. He turns to you and jokingly says,	Once inside the kitchen, you head to the kitchen sink. Your friend turns to you and says,
TARGET UTTERANCE		
‘Can you pour me some water?’		

Note: In each item, ‘Context A’ and ‘Target Utterance’ stayed the same, but context B was manipulated to produce the three experimental conditions.

There were 12 target utterances that were the same across each of the three conditions. The number of words and characters in each target utterance were based on Marocchini and Domaneschi’s (2022) study. All target utterances started with the words

‘Can you’ and were five to seven words in length (23 to 27 characters including spaces). Each utterance was composed of (1) the words ‘Can you’, (2) a transitive verb, and (3) a noun or pronoun. In addition, key pieces of information regarding the speaker’s perspective were highlighted in bold to make the information more salient and reduce the load on working memory. In this task, there were three control items that mirrored the structure of the target items. The crucial difference was that in the control items, the target utterance did not have a ‘Can you’ form (e.g., ‘How are you planning to get there?’).

To monitor whether participants properly understood the task instructions, four manipulation checks were included in the study. The manipulation checks gave examples of the definition of each condition, and the participant had to select the correct answer (e.g., If a person uses the phrase ‘Can you pass the salt?’ to mean ‘Are you physically able to pick up the salt and pass it to me?’, this would be an example of...). All materials were designed using the software platform Gorilla.

5.2.1 Norming study for the Pragmatic Language Task

Norming study aims

A norming study was performed to test and select the materials for the Pragmatic Language Task. In this paradigm, reading times were measured to characterise how social pragmatic utterances (ironic comments) are processed in comparison to the two control conditions (indirect request; literal utterance). It is imperative that the manipulated contexts in each condition lead to the correct interpretation of the target utterance. Thus, the aim of the norming study was to test how often the correct interpretation of the target utterance was inferred for each condition in each item. Marocchini and Domaneschi (2022) excluded participants who did not accurately identify the intended meaning of the target item at least

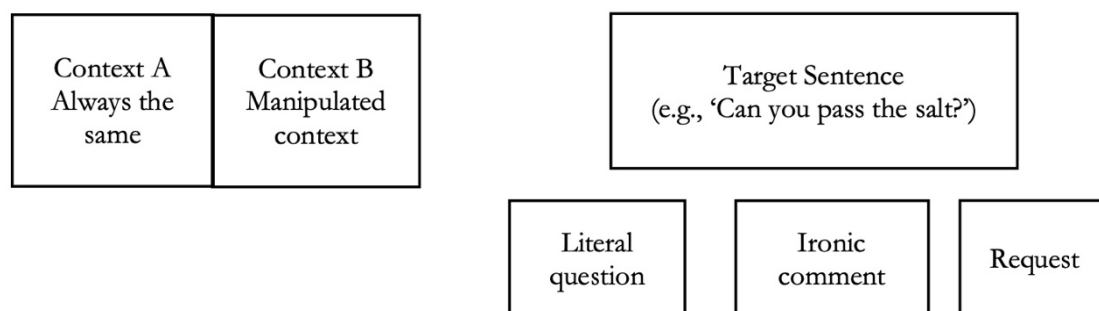
75% of the time. The reason that a high accuracy cut-off of 75% is needed is that the dependent variable in the main experiment is reading times, not accuracy. The dependent variable in the main experiment measures the amount of time it takes for a reader to process a specific type of pragmatic utterance (ironic comment, literal question, conventionalised indirect request). Therefore, to maintain the validity of the experiment, it is essential that there is a high level of certainty in what type of utterance is being measured. A potential downside to this is that it may mean that only the data of participants who are highly skilled at recognising different types of pragmatic utterances are retained. This may not accurately reflect how difficult it is to interpret pragmatic language in text in the general population. However, as the time it takes to process pragmatic utterances is the measured variable in this experiment, it is important to be sure which type of utterance is being measured. Therefore, narrative materials were selected for inclusion in the main study when they led to the correct interpretation of the target utterance more than 75% of the time following Marocchini and Domaneschi (2022).

Norming Study 1

The first, small-scale norming study used convenience sampling to recruit 11 participants (six females, five males), with a mean age of 40.7 ($SD = 4.9$). Here, 16 of the narratives from Trott and Bergen's (2020) study were adapted to create the three conditions (ironic comment, literal question, conventionalised indirect requests) for each item. Thus, the 16 narratives produced 48 conditions. Participants saw all conditions for each narrative. After reading each narrative, they completed a judgement task in which they selected whether the utterance was intended as an ironic comment, a literal question, or a conventionalised indirect request (see Figure 5.2).

Figure 5. 2

Presentation Sequence for Each Item in the Norming Study Stimulus Set



Note: The narrative (Context A and Context B) was presented on the same screen. After reading the narrative, the participants saw the target utterance on a separate screen below, which had three buttons. The participant indicated whether they thought the utterance was a literal question, an ironic comment, or a request by pressing one of the buttons.

The results of this study revealed that a manipulated context in the ironic comment condition and the conventionalised indirect request condition regularly led to a correct interpretation of the target utterance (see Table 5.4). However, in the non-directive literal condition, participants regularly mistook the utterance as a conventionalised indirect request.

Table 5. 4

*Total Number of Responses That Led to the Correct Interpretation of the Target Utterance;
11 Participants in Total*

Narrative Number	CONDITION		
	Literal Question	Ironie Comment	Request
	No. of correct responses out of 11 (% correct)	No. of correct responses out of 11 (% correct)	No. of incorrect responses out of 11 (% correct)
1	6 (55%)	7 (64%)	10 (91%)
2	6 (55%)	11 (100%)	11 (100%)
3	9 (82%)	11 (100%)	10 (91%)
4	N/A	11 (100%)	5 (45%)
5	6 (55%)	9 (82%)	9 (82%)
6	11 (100%)	10 (91%)	11 (100%)
7	6 (55%)	11 (100%)	9 (82%)
8	5 (45%)	11 (100%)	9 (82%)
9	6 (55%)	11 (100%)	11 (100%)
10	4 (36%)	8 (71%)	N/A
11	11 (100%)	9 (82%)	2 (18%)
12	2 (18%)	9 (82%)	10 (91%)
13	10 (91%)	9 (82%)	11 (100%)
14	7 (64%)	11 (100%)	10 (91%)
15	8 (71%)	11 (100%)	11 (100%)
16	11 (100%)	10 (91%)	11 (100%)
Total no. over 25% incorrect	10	2	3

Note: The figures highlighted in red show the total number of responses that were correct less than 75% of the time. N/A refers to conditions where no results were collected due to technical problems.

After this analysis of the results, nine out of the 16 narrative items were excluded (items 1, 4, 5, 7, 8, 9, 10, 11, 12) as the narratives' plots were deemed unsuitable to differentiate the three conditions. This left seven narrative items that regularly led to the

correct interpretation of the target utterance in each condition. The elements of these stories were analysed to identify elements that made the correct interpretation of target utterances more salient. After this analysis, eight new narrative items were created. This meant that there were 15 narrative items that could be tested for suitability in the norming study 2, with the aim of retaining 12 for the main study.

Norming Study 2

In this second norming study, 60 participants whose ages ranged from 19–64 ($M = 39.03$, $SD = 11.54$) were recruited via the recruitment platform Prolific. The study lasted approximately 15 minutes, and participants were paid £2.00 for taking part. Participants were excluded if they were not native English speakers or if they had any language disorders. Fourteen participants were excluded; nine of these failed more than one of the manipulation checks, and five completed the study in less than half of the time available, and their data were considered unusable.

In this study, there were 15 narratives, each presented under three conditions. Three counterbalanced lists were created in such a way that each participant saw one condition from each narrative item and an equal number of each condition; this meant that each participant saw five ironic comment conditions, five conventionalised indirect request conditions, and five non-directive literal conditions. Table 5.5 presents the results of the study, including how often the correct interpretation of the target utterance was inferred. The highest number of incorrect interpretations of the target utterance occurred in the literal question condition, and accuracy frequencies were correct less than 75% of the time in six narratives. In contrast, the accuracy frequencies were correct less than 75% of the time in only three of the ironic comment and conventionalised indirect request conditions.

Table 5. 5

This Describes the Total Number of Responses that Led to the Correct Interpretation of the Target Utterance.

Narrative Number	CONDITION		
	Literal Question	Ironic Comment	Request
	No. of correct responses out of 20 (% correct)	No. of correct responses out of 18 (% correct)	No. of correct responses out of 22 (% correct)
1	13 (65%)	16 (89%)	22 (100%)
2	11 (55%)	16 (89%)	16 (73%)
3	18 (90%)	17 (94%)	22 (100%)
4	12 (60%)	18 (100%)	19 (86%)
5	4 (20%)	17 (94%)	21 (95%)
6	19 (95%)	15 (83%)	22 (100%)
7	19 (95%)	18 (100%)	20 (91%)
8	19 (95%)	11 (61%)	20 (91%)
9	20 (100%)	18 (100%)	21 (95%)
10	14 (70%)	18 (100%)	21 (95%)
11	13 (65%)	13 (72%)	21 (95%)
12	17 (85%)	17 (94%)	21 (95%)
13	19 (95%)	16 (89%)	14 (64%)
14	20 (100%)	15 (83%)	16 (73%)
15	15 (75%)	12 (67%)	22 (100%)
Total no. over 25% incorrect	6	3	3

Note: The figures highlighted in red show the narrative items that had a total number of responses that were correct less than 75% of the time.

The results reveal that in two items, accuracy frequencies were correct less than 75% of the time in more than one condition (item 2 and item 11); for this reason, these two narrative items were excluded from the study. Narrative item 5 had a much lower accuracy

frequency rate in the non-directive literal condition (20%) than the other narratives, so this item was also excluded.

This left 12 narratives, the total needed for Study 4. Of these, in the literal question condition, three of the narrative items (item 1, item 4, item 10) contained accuracy frequencies that were correct less than 75% of the time. To make the intended literal meaning more salient in these narratives so they were suitable for the main study, several changes were made. Firstly, the text was altered to mirror the language used in the example items (e.g., not able to...) so the participants could more easily recognise the intended literal meaning. Secondly, additional key pieces of information were bolded. Thirdly, the final words before the question were altered to make it more obvious that the speaker was questioning whether the act could be fulfilled (see example below):

A partial excerpt of the literal question condition before editing:

CONTEXT: One afternoon, a regular elderly customer enters the store. Just as he approaches the counter, you both hear a shout from the back room: ‘Hey, we’re running low on chocolate chips for the muffins, so don't offer any more muffins to the customers.’ **The customer looks agitated when he hears this.** He politely asks,

UTTERANCE: ‘Can you get me a muffin?’

Partial excerpt of the literal question condition after editing:

CONTEXT: One afternoon, a regular elderly customer enters the store. Just as he approaches the counter, you both hear a shout from the back room: ‘Hey, we’re **running low** on chocolate chips for the muffins, so don't offer any more muffins as we may **not be able to make any more** for the customers.’ **The customer looks agitated when he hears this.** He **inquiringly** asks,

UTTERANCE: ‘Can you get me a muffin?’

In the ironic comment condition, two of the narrative items (item 8, item 15) contained accuracy frequencies that were correct less than 75% of the time. To make the intended ironic meaning more salient in these narratives so they were suitable for the main study, several changes were made. Firstly, the text was simplified, and descriptive details were cut to make the intended ironic interpretation more salient. Secondly, the vocabulary was altered to make it clearer at the end of the narrative that the comment was intended to amuse (see example below):

Partial excerpt of ironic comment condition before editing:

CONTEXT: When you get to the car you open the boot and find that it is full of clothes that you and Olivia were going to take to a charity shop. As you and Olivia stare at the open boot thinking about how to fit it all in, Aili comes up behind you with a couple of bags in each hand. As he approaches, **Aili sees that the boot is full of clothes, and that it will be impossible to fit the shopping in.** Aili looks at the full trolley and then at the bags in his hand. At this point, Aili looks at you and shakes his head at you. With a large grin he says,

UTTERANCE: ‘Can you fit this in the boot?’

Partial excerpt of ironic comment condition after editing:

CONTEXT: When you get to the car you open the boot and find that it is full of clothes that you and Olivia were going to take to a charity shop. You both stare at the open boot. Then, Aili comes up behind you with a couple of bags in each hand. As he approaches, **Aili sees that the boot is full of clothes and that it will be impossible to fit the shopping in.** At this point, you all start laughing. With a large grin Aili says,

UTTERANCE: ‘Can you fit this in the boot?’

Finally, in the conventionalised indirect request condition, two of the narrative items (item 13, item 14) contained accuracy frequencies that were correct less than 75% of the time.

To make the intended indirect request interpretation more salient in these narratives so they were suitable for the main study, several changes were made. In particular, the text was simplified, and descriptive details were cut to make the intended indirect request interpretation more salient.

Partial excerpt of conventionalised indirect request condition before editing:

CONTEXT: Unfortunately, you have claustrophobia, so you can't go into enclosed spaces like lifts or you will start to panic. You take the stairs up and see that Oliver already has a lot of stuff ready to take down. You grab a couple of bags and tell him that you'd prefer to use the stairs. Oliver, **who does not know you have claustrophobia**, looks surprised at this. He explains that using the stairs will make the move much longer and he only has the removal van for a couple of hours, so he says,

UTTERANCE: 'Can you take the lift?'

Partial excerpt of the conventionalised indirect request condition after editing:

CONTEXT: Unfortunately, you have claustrophobia, so you can't go into enclosed spaces like lifts or you will start to panic. You take the stairs up, which is very tiring. When you finally get to his floor, you see that Oliver already has a lot of stuff ready to take down. As you leave the flat, Oliver, **who does not know you have claustrophobia** says,

UTTERANCE: 'Can you take the lift?'

Mentalising and control tasks

None of the predictor (mentalising tasks) or control tasks (ART 2, Digit Span) were edited or changed.

Strange Stories – See Chapter 2, Section 2.2, pp 11.

Silent Films Task

The Silent Films task (Devine & Hughes, 2016) is a visual measure of mentalising. In this task, the participant watches five clips from a silent film: Harold Lloyd's classic silent comedy *Safety Last* (Roach et al., 1923). After watching each film clip, the participants answered a single mental state question about the protagonist's behaviour. There are six questions in total. To successfully answer the questions, the participant must have an understanding of deception, misunderstanding, and false belief. Accuracy was rated on a 0–2 scale by two researchers: 0 points for an incorrect answer, 1 point for a partially correct answer, and 2 points for a fully correct answer. The maximum score is, therefore, 16 points.

Adult Reading Task 2 (ART-2) – See Chapter 2, Section 2.2, pp. 12.

Digit Span Task – See Chapter 4, Section 4.2, pp 80.

5.1.2 Design

This current study aimed to characterise the way in which social pragmatic utterances (ironic comments) are processed in comparison to two control conditions, conventionalised indirect requests and non-directive literal questions, during the deployment of pragmatic inference in passage-length narratives. A second aim was to investigate how mentalising supports each type of utterance.

The Pragmatic Language Task involved a self-paced reading paradigm with a 1 x 3 within-participants design, with the condition (ironic comment, literal question, conventionalised indirect request) as the independent variable. For each of the three conditions in each narrative item, the target utterance was the same (see Table 5.4). The dependent variables were (1) the time taken to read each of the three regions of the target utterance (regional reading time) and (2) the time taken to read the whole target utterance (global reading time). After reading the narrative and utterance in each item, participants performed a judgement task in which they were asked to identify whether the utterance was intended to be an ironic comment, literal question, or conventionalised indirect request. Participants who scored below 75% accuracy on the judgement task had their results removed from the analysis. In this task, three counterbalanced lists were created so that each participant read only one condition from each of the 12 target items. The target stories and controls were presented in the same order.

To investigate the involvement of mentalising during the deployment of pragmatic inference in ironic comments and the two control conditions, the Strange Stories and Silent Films Task were used. If strong correlations between the two mentalising tasks were found, a latent variable factor would be created from the scores of these two tasks. However, recent research suggests there is often little coherence between different mentalising tasks as they measure different aspects of mentalising and associated abilities (Warnell & Redcay, 2019).

To control for shared verbal properties in the Strange Stories and the Pragmatic Inference Task, Section 2 of ART-2 was included. Additionally, as working memory was found to play a prominent role in deployment in Study 3, a forward and backwards digit span task was included to control for working memory.

5.1.3 Procedure

The study took approximately 45 minutes to complete, and all tasks were designed and run on the test-building software Gorilla. Testing took place online via the recruitment platforms Prolific and Sona. Participants on Prolific received £5.70 for their participation, and participants on Sona were students from York St. John University who received course credits. Participants first read the participant information guide and gave consent, and then they were asked for demographic details regarding age, ethnicity, occupation, and gender. The participants were asked to complete the task in a quiet room in which they could concentrate, and in Gorilla the settings only allowed participants to complete the experiment on a computer (rather than a phone or tablet). The participants then completed the Pragmatic Language Task (pragmatic language and reading comprehension), the Digit Span task (working memory), the Strange Stories (verbal mentalising task), the Silent Films (visual mentalising task), and finally the ART-2 (reading comprehension).

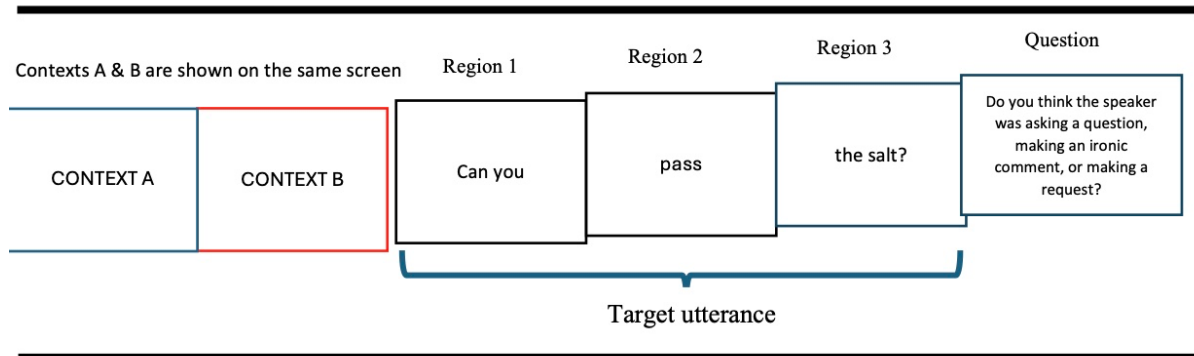
Before starting the Pragmatic Language Task, participants read instructions that described the characteristics of ironic comments, literal questions, and conventionalised indirect requests. Next, the participants completed a practice round, in which they saw all three conditions of a practice item. In the practice round, they first read a narrative and an utterance, then they decided whether it was an ironic comment, literal question, or conventionalised indirect request. After making their decision, they were given the correct

answer and an explanation as to why the answer was correct. Participants were instructed to read at their own pace and to click the mouse button to continue as soon as they finished reading each section; before each item, the participants saw an instruction screen that reiterated the instructions. Reading times were measured from the time the screen appeared to the time the mouse button was pressed. Reading times were recorded in milliseconds. After reading the narrative and target utterance for each item, they performed the judgement task.

In each item, the narrative, including contexts A and B, was presented together on a single screen, whereas the target utterance was presented over three separate screens (see Figure 5.3). This enabled measurement of the reading times for each of the three regions of the target utterance (Target 1, Target 2, Target 3) as well as the overall reading time for all three target regions. To progress from one screen to another, the participant clicked the mouse button. After the participants read the narrative and target utterance, they were asked to judge whether the speaker was asking a question, making an ironic comment, or making a request.

Figure 5. 3

Example Screenshots of the Indirect Request Condition in Study 4.



Note: Only the context 2 screen was manipulated.

5.1.4 Statistical analysis

Analyses were conducted in R using linear mixed models (LMMs), Lme4 (Bates et al., 2015), and the LmerTest package. Tukey contrasts were used for post-hoc comparisons using the emmeans package. Analyses were conducted in two stages. Firstly, the differences in reading times for each condition were compared. Next, whole sentence reading times and each of the three regional reading times for each of the three conditions (ironic comments, literal questions, and conventionalised indirect requests) were analysed, with reading times as the dependent variable and fixed effects of the condition as well as random intercepts as by subject random effects. Thus, in the first stage of the analysis, four LMMs were built: three for each sentence region and one for whole sentence reading times.

The second stage of the analysis investigated whether the two mentalising measures predict performance on the Pragmatic Language Task when controlling for working memory

and reading comprehension. There was a small significant positive correlation between the Strange Stories and the Silent Films Task (.35, $p < .001$), but this was smaller than expected. The Silent films Task was designed as a companion task to the Strange Stories, measuring the same mentalising aspects but drawing less on verbal skills (Devine and Hughes, 2013). Therefore, we would have expected higher correlations between the two tasks; as such, we did not create a mentalising latent variable. Therefore, an LMM was run for each mentalising predictor separately. At this second stage of analysis, an LMM was run for whole- sentence reading times and for each region of the target sentence, with reading time as the dependent variable, mentalising as the independent variable, condition and the two control measures, reading comprehension and working memory as fixed effects, and an interaction between the fixed effects; random intercepts were also included. This meant that there were another eight LMMs to analyse the relationship of mentalising with processing times in each condition.

The main advantage of using linear mixed models (LMMs) to measure reading times is that LMMs control for by-subject effects, so they control for individual differences between participants (Baayen et al.2008). In contrast, an ANOVA assumes that all variability is captured at the group level, which ignores individual differences. Furthermore, because LMMs incorporate individual differences as random effects, they reduce error variance and increase statistical power compared to ANOVA (Barr, Levy, Scheepers, & Tily, 2013).

5.3 Results

Judgement task results

There was high accuracy frequency on the judgement task (86.30% correct), which asked participants to identify whether each utterance was an ironic comment, literal question, or a conventionalised indirect request.

5.3.1 Comprehension results on the Pragmatic Language Task

The descriptive data for the 90 participants in the Pragmatic Language Task are available in Table 5.6. Descriptive statistics for the mentalising measures and controls are presented in Table 5.7.

Table 5. 6

Mean Reading Time Scores for Each Region of the Target Utterance and Whole Sentence Reading Time (Total)

PRAGMATIC LANGUAGE TASK				
Condition	Region 1 Mean (SD)	Region 2 Mean (SD)	Region 3 Mean (SD)	Total Mean (SD)
Ironic Comment	709.57 (290.53)	474.06 (186.76)	602.93 (300.44)	1,786.56 (616.31)
	<i>R</i> = 737.59-649.57	<i>R</i> = 532.69-367.84	<i>R</i> =1066.03-284.24	<i>R</i> =2284.28-1389.65
Literal Question	726.02 (295.68)	501.81 (214.78)	671.59 (588.67)	1,899.42 (812.71)
	<i>R</i> = 782.56-484.88	<i>R</i> = 866.09-367.84	<i>R</i> = 715.90-568.48	<i>R</i> =2364.56-1421.2
Indirect Request	710.67 (257.02)	474.72 (175.21)	599.13 (284.10)	1,784.51 (573.72)
	<i>R</i> = 2285.84-316.36	<i>R</i> = 1282.90-100.37	<i>R</i> = 2583.15-49.99	<i>R</i> = 4266.15-800.59

* *R* – is the range of scores
Descriptive Statistics

Table 5. 7*Descriptive Statistics*

Task	Mean (SD)	Range	Max. Score
Strange Stories	12.97 (2.11)	7–16	16
Digit Span	5.49 (1.66)	2–9	10
Intentionality	12.26 (3.07)	3–19	20
ToM composite	15.53 (4.69)	3–27	28
<i>Frith-Happé - GD</i>			
Appropriateness	4.99 (2.27)	0–8	8
Intentionality	8.05 (3.40)	0–14	20
GD composite	14.82 (2.35)	8–20	28

Note. Frith Happé – ToM = Frith-Happé mentalising animations, Frith-Happé GD = Frith-Happé goal-directed animations. The ToM composite score was produced by combining the Frith-Happé appropriateness and intentionality scores, and this was used as the final score on the Frith-Happé animations.

Table 5. 8*Mean Scores for mentalising variables and controls*

Task	Mean (SD)	Range	Max. Score
Strange Stories	10.95 (2.28)	3-15	15
Silent Films Task	8.54 (1.67)	3-11	11
Digit Span	5.48 (1.60)	2-9	9
ART 2	5.11 (2.05)	1-9	9

Correlation between mentalising variables and controls

The Strange Stories and the Silent Films Task were found to be moderately correlated, $r(90) = .35, p < .001$. There were also small correlations between the Strange Stories and reading comprehension, $r(90) = .23, p < .001$, and the Strange Stories and working memory, $r(90) = .16, p < .001$.

Table 5. 9

Correlation Between Mentalising Variables and Control Variables

	1.	2.	3.	4.
1. Silent Films Task				
2. Strange Stories	.35**			
3. Digit Span Task	.00	.16**		
4. Reading comprehension (ART 2)	.01	.23**	.35**	

* $p < .05$, ** $p < .001$

Note: This shows the correlations between the mentalising variables of Strange Stories and Silent Films Task, and the control variables of reading comprehension (ART-2) and working memory (digit span).

Table 5. 10

Correlations Between Mentalising and Control Variables and the Reading Times on the Pragmatic Language Task

Variable	Digit Span	ART-2	Strange Stories	Silent Films
Irony comments				
Region 1	-.140**	-.019	-.030	-.003
Region 2	.011	.072	.033	.020
Region 3	.027	.039	.020	-.026
Whole Sentence	-.050	.032	.006	-.008
Indirect Requests				
Region 1	-.047	-.047	.000	.041
Region 2	-.001	.093	-.007	-.037
Region 3	.031	.038	-.038	-.064
Whole Sentence	-.006	.026	-.021	-.025
Literal Questions				
Region 1	-.039	-.002	.028	-.024
Region 2	-.010	.035	.005	-.057
Region 3	.036	-.014	-.031	-.021
Whole Sentence	.010	-.002	-.011	-.039

* $p < .05$. ** $p < .01$.

Note. There was a significant correlation (-.140**) between the Digit Span Task and the Irony comment condition (region 1) of the Pragmatic Language Task. There were no other significant correlations.

The LMM results for whole sentence reading times showed a main effect of the condition ($F[2, 991.86] = 6.83; p < .001$). When examining the pairwise comparisons, it becomes apparent that ironic comment utterances were read significantly faster than literal question utterances ($b = -112.86; SE = 35.6; p = .005$), but there was a negligible difference

in reading times between ironic comments and conventionalised indirect request utterances ($b = 2.05$, $SE = 34.6$, $p = .998$). Literal question utterances were read significantly slower than conventionalised indirect request utterances ($b = 114.90$, $SE = 35.6$, $p = .004$).

Next, differences in reading times across the conditions were examined for the three regions: target region 1 (start of the utterance), target region 2 (middle part of the utterance), and target region 3 (end of the utterance).

Target sentence region 1

The LMM results for target region 1 reading times found a non-significant main effect for the condition ($F[2, 991.86] = 0.71$; $p = .490$).

Target sentence region 2

The LMM results for target region 2 reading times found a significant main effect for the condition ($F[2, 991.93] = 4.64$; $p < .001$). Examining the pairwise comparisons, ironic comment utterances were read significantly faster than literal question utterances ($b = -27.75$; $SE = 10.4$; $p = .021$), but there was again a negligible difference in reading times between the ironic comment and conventionalised indirect request utterances ($b = -0.66$, $SE = 10.4$, $p = .998$). Literal question utterances were read significantly slower than conventionalised indirect request utterances ($b = 27.01$, $SE = 10.4$, $p = .025$).

Target sentence region 3

The LMM results for target region 3 reading times revealed a significant main effect of condition ($F[2, 991.73] = 4.99; p = .007$). Per the pairwise comparisons, ironic comment utterances were read significantly faster than literal question utterances ($b = 11.07$, $SE = 11.66$, $DF = 989.30$, $t = 0.95$, $p = .34$), but there was little difference in reading times between ironic comment and conventionalised indirect request utterances ($b = 3.8$, $SE = 25.8$, $p = .998$). Literal question utterances were read significantly slower than conventionalised indirect request utterances ($b = 72.5$, $SE = 25.8$, $p = .014$).

5.3.3 Analysis of predictors

Effect of mentalising (Strange Stories)

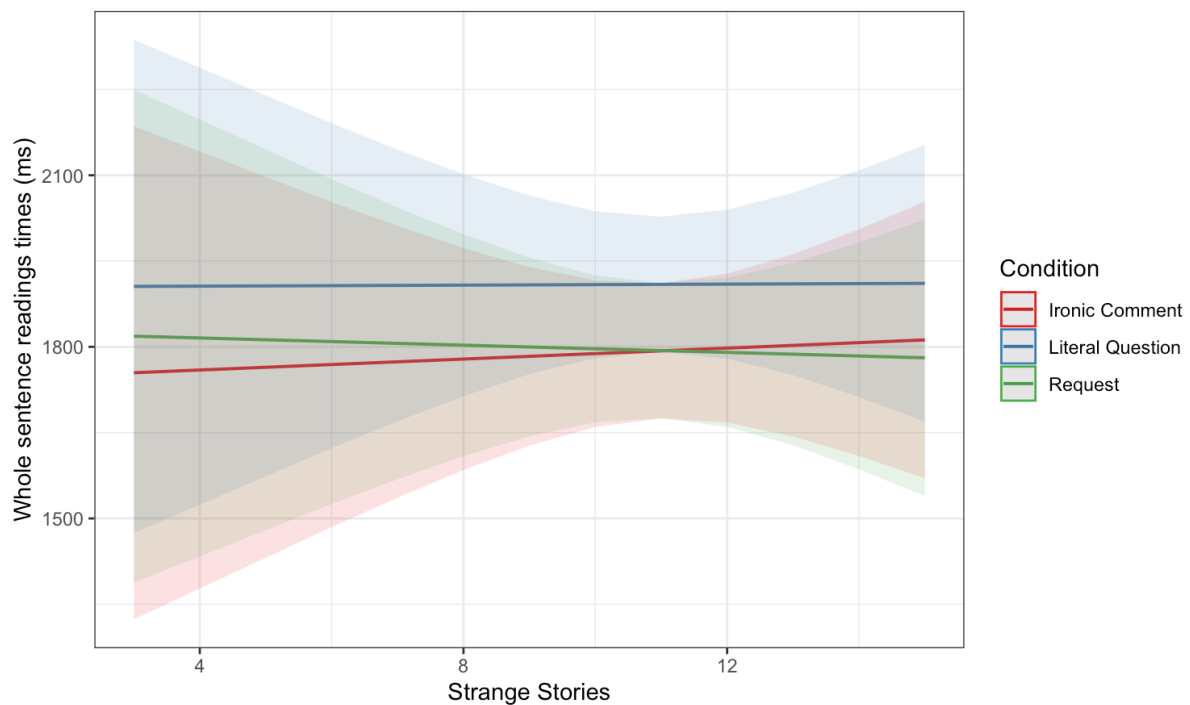
Whole sentence reading times

The results of the LMM revealed no significant main effect of mentalising on whole sentence reading times when using the Strange Stories as a predictor ($b = -35.14$, $SE = 25.45$; $DF = 126.66$; $t = -1.38$; $p = .17$). There were also no significant interactions found between the scores on the Strange Stories and whole sentence reading times in the ironic comment condition ($b = 5.61$, $SE = 16.04$, $DF = 986.02$, $t = 0.35$, $p = .73$) or the

conventionalised indirect request condition ($b = -1.43$; $SE = 16.04$; $DF = 986.02$; $t = -0.09$; $p = .93$).

Figure 5. 4

The relationship between the Strange Stories and whole sentence reading times in each condition



Note. There were no significant interactions found between the Strange Stories and whole sentence reading times in any of the three conditions.

Target Region 1

The results of the LMM found no significant main effect of mentalising on region 1 reading times when using the Strange Stories as a predictor ($b = -8.78$, $SE = 10.42$; $DF =$

128.50; $t = -0.84$; $p = .40$). There were also no significant interactions found between the scores on the Strange Stories and region 1 reading times in the ironic comment condition ($b = -7.70$, $SE = 6.90$, $DF = 986.08$, $t = -1.11$, $p = .27$) or the conventionalised indirect request condition ($b = -3.78$; $SE = 6.94$; $DF = 986.08$; $t = -0.54$; $p = .59$).

Target Region 2

The results of the LMM indicated no significant main effect of mentalising on region 2 reading times when using the Strange Stories as a predictor ($b = -4.90$, $SE = 7.11$; $DF = 131.09$; $t = -0.69$; $p = .49$). There were also no significant interactions between the scores on the Strange Stories and region 2 reading times in the ironic comment condition ($b = 2.24$, $SE = 4.69$, $DF = 988.13$, $t = 0.48$, $p = .63$) or the conventionalised indirect request condition ($b = -1.08$; $SE = 4.69$; $DF = 988.13$; $t = -0.23$; $p = .82$).

Target Region 3

The results of the LMM found no significant main effect of mentalising on region 3 reading times when using the Strange Stories as a predictor ($b = -12.73$, $SE = 13.79$; $DF = 154.75$; $t = -0.92$; $p = .36$). There were also no significant interactions between the scores on the Strange Stories and region 3 reading times in the ironic comment condition ($b = 11.07$, $SE = 11.66$, $DF = 989.30$, $t = 0.95$, $p = .34$) or the conventionalised indirect request condition ($b = 3.41$; $SE = 11.66$; $DF = 989.30$; $t = 0.29$; $p = .77$).

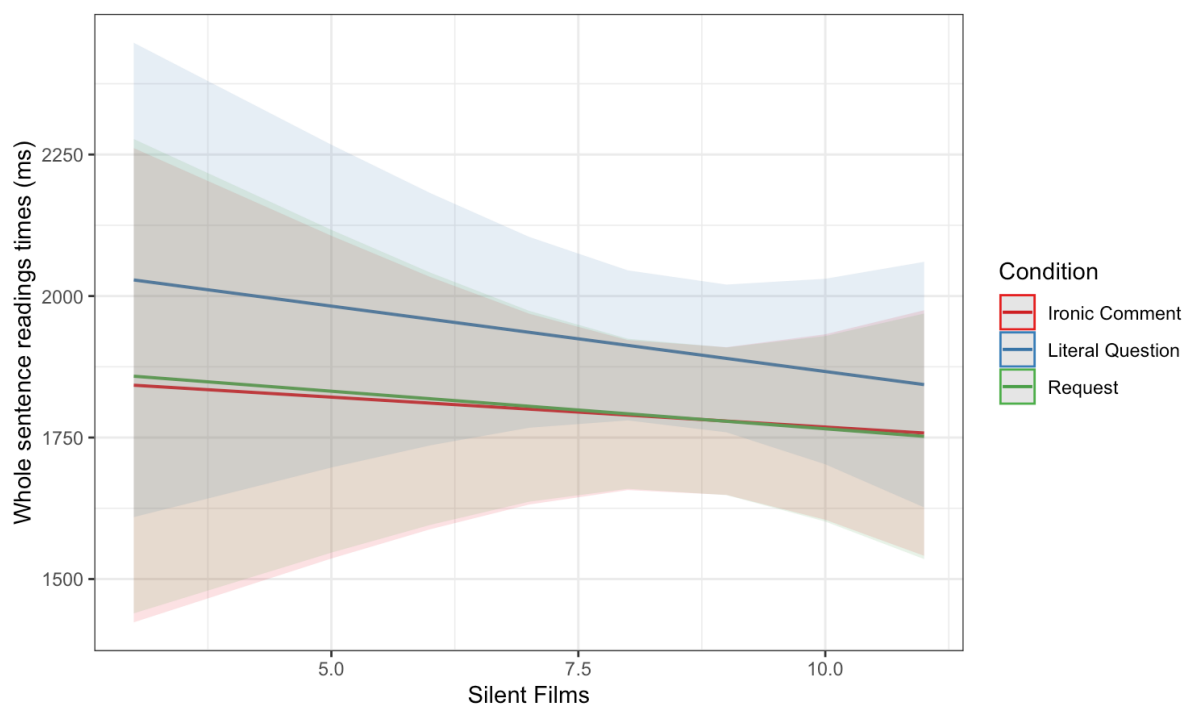
Effects of mentalising (Silent Films)

Whole sentence reading times

The results of the LMM did not reveal a significant main effect of mentalising on whole sentence reading times when using Silent Films as a predictor ($b = 56.01$, $SE = 31.53$; $DF = 216.50$; $t = 1.78$; $p = .08$). Additionally, there were no significant interactions between the scores on the Silent Films and whole sentence reading times in the ironic comment condition ($b = 16.57$, $SE = 21.70$, $DF = 981.43$, $t = 0.76$, $p = .45$) or the conventionalised indirect request condition ($b = 10.73$; $SE = 21.70$; $DF = 981.43$; $t = 0.96$; $p = .62$).

Figure 5. 5

The relationship between the Silent Films Task and whole sentence reading times in each condition.



Note. There were no significant interactions found between the Silent Films and whole sentence reading times in any of the three conditions.

Target Region 1

The results of the LMM found no significant main effect of mentalising on region 1 reading times when using Silent Films as a predictor ($b = 17.64$, $SE = 12.99$; $DF = 212.68$; $t = 1.36$; $p = .18$). There were also no significant interactions between the scores on the Silent Films and region 1 reading times in the ironic comment condition ($b = 3.95$, $SE = 9.40$, $DF = 985.01$, $t = 0.42$, $p = .68$) or the conventionalised indirect request condition ($b = 10.82$; $SE = 9.40$; $DF = 985.01$; $t = 1.15$; $p = .25$).

Target Region 2

The results of the LMM revealed no significant main effect of mentalising on region 2 reading times when using Silent Films as a predictor ($b = 7.11$, $SE = 8.86$; $DF = 216.36$; $t = 0.80$; $p = .42$). There were also no significant interactions between the scores on the Silent Films and region 2 reading times in the ironic comment condition ($b = 9.82$, $SE = 6.36$, $DF = 985.94$, $t = 1.55$, $p = .12$) or the conventionalised indirect request condition ($b = -3.51$; $SE = 6.36$; $DF = 985.94$; $t = 0.55$; $p = .58$).

Target Region 3

The results of the LMM found no significant main effect of mentalising on region 3 reading times when using Silent Films as a predictor ($b = 9.57$, $SE = 17.99$; $DF = 206.61$; $t = 0.53$; $p = .60$). There were also no significant interactions between the scores on the Silent Films and region 3 reading times in the ironic comment condition ($b = 2.73$, $SE = 15.82$, $DF = 987.33$, $t = 0.177$, $p = .86$) or the conventionalised indirect request condition ($b = -3.60$; $SE = 15.82$; $DF = 987.33$; $t = -0.23$; $p = .82$).

5.4 Discussion

The first aim of this study was to characterise the way in which individuals process a form of social pragmatic language (ironic comments) across a whole sentence in comparison to a linguistic pragmatic form (conventionalised indirect requests) and literal questions. To directly compare these three different types of utterances, three conditions were created in which the context of a narrative was manipulated to produce three different interpretations of an utterance (e.g., ‘Can you pass the salt?’). By doing this, it was possible to directly compare the reading times of the utterance in each condition at (1) the local level during three separate parts of the utterance, and (2) the global level across the whole sentence.

The second aim was to investigate whether mentalising supported the processing of ironic comments. To do this, two mentalising tasks were used: Strange Stories (a verbal mentalising task) and Silent Films (a visual mentalising task). If a relationship between the mentalising predictors and the Pragmatic Language Task did exist, LMMs would be run to see if this relationship was still significant when controlling for working memory and reading comprehension.

5.4.1 Whole sentence and regional reading times: Pragmatic Language Task

RQ1. How are ironic comments (social pragmatic) processed in comparison to conventionalised indirect requests (linguistic pragmatic) and non-directive literal questions at the local and global levels?

Summary of the predictions and findings – whole sentence level

When interpreting ironic comments, the reader must consider the context surrounding the utterance and the speaker's intentions to infer the intended meaning of the utterance as the literal meaning of the utterance is incongruous with the intended meaning. Hence, when interpreting ironic comments, the reader must integrate information from the context with the literal meaning of the ironic utterance to form a meta-representation of the whole sentence, which may recruit mentalising (Happé, 1993; Wilson, 2009). In contrast, when processing literal questions, the reader must parse the sentence word by word, and meaning is only gained when the whole sentence is read. Literal statements are processed more rapidly than ironic comments (Marocchini & Domaneschi, 2022). On the other hand, the intended directive meaning of conventionalised indirect requests is signalled by linguistic cues at the start of the sentence (e.g., 'Can you'). Therefore, the directive meaning of the utterance is interpreted before the whole sentence is read (Reboul, 2004). Interpreting the meaning of a conventionalised indirect request is thought to be even more rapid than interpreting literal questions because readers have been so frequently exposed to the directive interpretation of 'Can you' utterances that the directive interpretation is often derived more than the non-directive literal meaning (Gibbs, 1983; Morgan, 1978). It was therefore predicted that there would be longer whole sentence processing times in the ironic comment condition than in the literal question and conventionalised indirect request conditions.

However, contrary to expectations, the results showed no significant difference in reading times globally across the whole sentence between the ironic comment condition and the conventionalised indirect request condition. In addition, the results revealed that reading

times across the whole sentence in the literal question condition were significantly more slowly than either conventionalised indirect requests or ironic comments. There was no significant difference in reading times between the ironic and conventionalised indirect request conditions.

Summary of findings – across each region of the utterance

At the regional level, given that ironic comment utterances need to be heard in full and integrated with contextual cues to derive the intended meaning, it was expected that the results would feature similar reading times across each region of the utterance, with longer reading times at the end of the utterance (region 3) that indicate greater processing at this point (i.e., wrap-up effects). It was expected that the same pattern of results would be observed in the literal question condition, because the sentence needs to be read in full to derive the intended meaning. However, as meta-representation is not necessary for the interpretation of literal comments, it was expected that faster reading times would be found in each region of the target utterance. In contrast, conventionalised indirect requests contain lexical cues that signal that a request is forthcoming. In the Pragmatic Language Task, the lexical cue words ('Can you') in region 1 should signal the directive meaning; as such, it was expected that increased processing and longer reading times would be found at this stage in comparison with the literal question condition because the intended meaning of literal questions is interpreted once the whole sentence has been processed. Slower reading times would therefore be expected in both the ironic comment condition and the conventionalised indirect request condition at region 1 than in the literal question condition. However, slower reading times were expected across regions 2 and 3 in the ironic comment condition than the conventionalised indirect request condition and literal question condition.

At the local level, the results also did not support the predictions. There were no significant differences between reading times across all three conditions locally in region 1. Again, the literal question condition was read significantly more slowly locally across regions 2 and 3 of the utterance than the other two conditions. Across regions 2 and 3, both the ironic comment and the conventionalized indirect request conditions were read faster than the literal question condition. There were no significant differences in reading time between the ironic and conventionalised indirect request conditions.

Summary of the findings

Local and global reading times were similar in the ironic comment condition and the conventionalised indirect request condition. However, the literal question condition was read significantly more slowly globally, across the whole sentence, locally, and in regions 2 and 3. To understand these results, it is necessary to examine the nature of the results in the control conditions and compare them to the ironic comment condition to characterise the way in which ironic comments have been processed in this experiment.

To characterise the way in which ironic comments are processed, they were compared to control utterances that are interpreted through lexical-based strategies and are not thought to involve meta-representation or recruit mentalising during processing. Conventionalised indirect requests contain lexical cues that signal the utterance's meaning before the whole sentence has been heard. When the reader encounters the lexical cue words 'Can you' at the start of the sentence (region 1), they recognise that the utterance is a request and so read the middle (region 2) and the end (region 3) of the utterance more rapidly as the intended function of the sentence has already been understood. Indeed, like Marocchini and Domaneschi (2022), longer reading times were found in region 1 and short reading times in

regions 2 and 3, indicating that the meaning of the utterance in this condition was comprehended after reading the lexical cue words 'Can you'.

The second comparison utterance was a non-directive, literal question. When studying the way in which conventionalised indirect requests were processed in comparison to literal questions, researchers focused on testing the serial model of processing (Clark & Lucy, 1975). This model suggests that when comprehending the meaning of an indirect request, the reader will derive the literal meaning of the utterance before inferring the indirect meaning of the utterance. However, this theory has not been supported by experimental studies. Indeed, research examining conventionalised requests suggests that the directive meaning of the utterance is accessed directly without the need for first computing the literal meaning. Gibbs (1983) not only demonstrated that the indirect meaning of an utterance is processed as fast as the literal meaning in 'Can you' sentences but also that people tend to comprehend the directive-indirect meaning even when the context surrounding an utterance is biased towards the non-directive literal meaning. These findings have been explained through the short-circuited implicature view, whereby people are more likely to derive the meaning they have heard most frequently in the past (Morgan, 1978). The results support these findings and show that literal questions were read significantly more slowly across the whole sentence, and specifically across regions 2 and 3, than 'Can you' conventionalised indirect requests. These results suggest that increased processing is needed to derive the correct interpretation in the literal question condition; this could be because when reading an utterance (e.g., 'Can you pass the salt?'), readers need to inhibit their tendency to derive the directive indirect interpretation of the utterance (e.g., 'Pass the salt') that they have heard more frequently. As individuals are so accustomed to deriving this meaning in everyday life, it may be more cognitively taxing to arrive at the literal meaning (e.g., 'Are you able to pass the salt?') even when the context of the narrative is

heavily biased towards this interpretation. Indeed, in the norming study, it was in the literal question condition that participants most often incorrectly identified the intended interpretation of the utterance, which was regularly mistaken as a conventionalised indirect request. When comparing reading times between literal questions and conventionalised indirect requests at region 1, no significant difference was found between literal questions and conventionalised indirect requests, but this is expected as readers spend longer on the lexical cue words ‘Can you’ that trigger meaning in conventionalised indirect requests. Thus, the pattern of processing utterances in the control conditions supports those found in previous research (Marocchini & Domaneschi, 2022).

Utterance processing in the ironic comment condition

The results of reading times in the ironic comment condition did not support the predictions or the pattern of results expected from previous research (Marocchini & Domaneschi, 2022). As ironic comments are thought to be understood through metarepresentation and to be more cognitively taxing than either literal utterances or conventionalised indirect requests, longer reading times were expected globally, across the whole sentence, and locally in each region of the ironic utterance. Given that the results show that ironic comments were read much faster than literal questions both globally and locally, one possible conclusion is that meta-representation was not needed to interpret the intended meaning of utterances in the ironic comment condition. A finding that supports this is that in the ironic comment condition, reading times were longer in region 1 than regions 2 and 3, and therefore there were no signs of wrap-up effects that would indicate integrative processing at the end of the sentence (Rayner et al., 2000). This suggests that the intended meaning of the utterance had already been derived before the whole sentence was read. This

finding is in keeping with theories of humour in spoken language that suggest we note the intent to amuse before processing the ironic utterances, which is signalled by cues such as mood and laughter that may have been primed in the text (Hocika & Gattis, 2008; Yus, 2003). In the Pragmatic Language Task used in this study, the speakers' intended meaning (to amuse) in the ironic comment condition may have been gathered from cues in the preceding narrative.

Andrés-Roqueta and Katos (2017) propose that the relationship between pragmatics, structural language, and mentalising is a dynamic process that is dependent on the specific properties of the interaction. Sometimes the meaning of a pragmatic utterance can be understood through the reader's own linguistic abilities, knowledge of pragmatic norms, and their egocentric point of view, without the need to represent the mental state of the speaker (O'Neil, 2012). It is possible that the irony used in the narratives can be inferred in this way. For example, below is an abbreviated excerpt from a narrative in the ironic comment condition.

CONTEXT: This morning, during a house meeting with you, Lisa, and Brian, you tetchily state that you have ended up tidying for Lisa for the last three weeks and refuse to do it again. Noting your irritation, Lisa volunteers to take the lead on cleaning the kitchen when she gets home from work this evening. **Brian seems satisfied with this plan.**

In the afternoon, while Lisa is still at work, you and Brian are standing in the kitchen doorway, gazing in. Brian turns to you and jokingly says,

UTTERANCE: 'Can you clean the kitchen?'

To understand why the utterance is ironic, we must be aware of the context: three people are in a house meeting to decide who should clean the house; ‘you’ refuse to clean, and so Lisa agrees to do it in the evening. If we assume that we must take the perspective of the speaker to interpret the meaning of the utterance, we must note that ‘Brian is satisfied with this plan’ (i.e., Lisa doing the cleaning), which is juxtaposed with the literal meaning of his utterance, ‘Can you clean the kitchen?’ However, it is also arguable that we can merely interpret the utterance by understanding the context and noting Brian’s intention to be funny through the words ‘jokingly says’, and therefore a meta-representation of the utterance may not be necessary. Participants may have been over-primed to expect an ironic utterance to follow the context.

However, if meta-representation is not necessary to comprehend ironic utterances, it may be expected that the reading times in this condition would be comparable to those in the literal question condition. Yet the reason that the utterance takes longer to process in the literal question condition may be because, as mentioned, ‘Can you’ utterances frequently occur in common English as conventionalised indirect requests; due to this, even in narratives biased towards the literal interpretation of the utterance, the reader must inhibit their tendency to take the directive indirect interpretation (see example of literal question condition below).

CONTEXT: This morning, during a house meeting with you, Lisa, and Brian, you state that you are too busy today to clean the house as you are working from home. So, Lisa volunteers to take the lead on cleaning the kitchen when she gets home from work this evening. **Brian seems satisfied with this plan.**

In the afternoon, while Lisa is still at the office, you take a break from work. While heading to the kitchen, you see Brian standing in the kitchen doorway, gazing in. Brian, **who knows you're busy and that you may not be able to spare any time, hopefully asks,**

UTTERANCE: 'Can you clean the kitchen?'

Another reason, deriving a literal meaning from the context is potentially harder than deriving an ironic interpretation of the utterance. Despite the bolding of key pieces of information that bias the interpretation of the utterance towards a literal interpretation, the reader may find it harder to derive the correct interpretation as the context is more ambiguous than in the ironic comment condition, which is reflected in the longer processing times. The longer reading times for utterances in the literal question condition highlight the influence of context in pragmatic language. Hocika (2014) states that 'Humour highlights a central aspect of pragmatics – that context and intentions are more important to interpreting pragmatic utterances, above and beyond literal meaning' (p.219). Clear contexts may be even more important to pragmatic language in narrative texts than in conversational spoken language. If breakdowns in conversation occur in spoken language, such as misunderstanding the intended meaning of an utterance, we can ask for clarity (e.g., 'Sorry, do you mean am I able to clean the kitchen, or do you want me to clean the kitchen?'). When reading, however, we can only re-read the text, and if the context is not clear, the reader may still struggle to derive the intended meaning of the utterance. Therefore, faster

reading times may have resulted from the ironic comment condition versus the literal condition as the context in the ironic condition was clearer. However, although the more ambiguous context could explain longer reading times in the literal condition, longer reading times would be expected in the ironic comment condition as individuals need to identify the speaker's intentions to interpret the utterance as ironic. The reason this was not found is that, firstly, the readers were heavily primed to expect an ironic utterance, and secondly the ways to identify a speaker's intentions in text are more rapid in comparison to spoken communication.

5.4.2 Does mentalising predict reading times in the ironic comment condition?

RQ2. To what extent does mentalising predict the deployment of the speaker's perceptual information for pragmatic inference in ironic comments in comparison to conventionalised indirect requests and non-directive literal statements?

Summary of the predictions and findings at the whole sentence and regional levels

The second aim of this study was to investigate the role of mentalising in the interpretation of social pragmatic language (ironic comments). As it was predicted that the interpretation of ironic utterances would be based on the whole content of an utterance and meta-representation, mentalising was anticipated to predict faster reading times across the whole sentence and each region of the target utterance. However, neither of the two mentalising variables (Strange Stories and Silent Films Task) predicted performance in the ironic comment or conventionalised indirect conditions in the Pragmatic Language Task.

The reason that mentalising did not predict performance in the ironic comment condition may be due to the nature of the Pragmatic Language Task used in this study. Mentalising is recruited flexibly during reading depending on the properties of the text. As discussed, the context of the ironic comment condition may have over-primed the participants to expect an ironic utterance to follow the story. Thus, the participants only needed to rely on their linguistic abilities and their own egocentric point of view to interpret the intended ironic function, rather than fully representing the speakers' perspectives (O'Neil, 2012). It is therefore possible that the properties of the text determine whether mentalising is recruited or not. In narrative texts where the context is primed with lexical cues to signal that an utterance is ironic, mentalising may not be needed to interpret the meaning of an ironic utterance. In contrast, it is possible that longer and more sophisticated texts that involve much greater character development (e.g., literary texts) may be more reliant on the reader's ability to contrast the perspectives of characters with the literal meaning of their utterances to interpret irony. In this case, mentalising may be recruited. Thus, both oral and written pragmatic language comprehension may involve dynamic relationships between mentalising, structural language, and pragmatic language that alters depending on the interaction (Andrés-Roqueta & Katos, 2017). When reading, this relationship may be dependent on the properties of the text. However, there is currently little research that has examined this area.

On the other hand, there are qualitative differences in the way that pragmatic language is processed in written language in comparison to spoken language, which suggests that readers are more reliant on coherent contextual information to understand the intentions behind spoken dialogue. To ensure smooth conversation during spoken dialogues, interlocutors must become sensitive to each other's perspectives. In spoken communication, individuals do this by using attention and memory to monitor signals that accompany

language and enhance meaning, such as vocal intonation, facial expressions, body movements, and a wide variety of forms of laughter (e.g., snigger, belly laugh; Apperley, 2018). In text, readers do not have access to many of these signals. For instance, the intonation that would accompany a literal question in the 'Can you' form (e.g., raising of pitch) in spoken language and aid interpretation of the utterance is not available in text. In passage-length written texts, the reader may be more reliant on understanding the context to interpret the speaker's spoken dialogue as ironic, since readers only have information from the text to decipher the speaker's intended meaning. Thus, the reader may be more reliant on descriptive details and linguistic cues (e.g., 'she said snidely') to interpret the intended meaning of an utterance. Accordingly, interpreting irony during reading may put more stress on readers' linguistic skills, and there may be less needed to fully represent the speakers' perspectives to interpret verbal irony. This partially explains why this study found that mentalising did not support the deployment of pragmatic inference.

Alternatively, it is possible that the reader does use Brian's perspective to interpret the meaning of the utterance but only needs to become sensitive to Brian's perspective through memory and attention rather than fully representing it (see example of ironic comment condition below).

CONTEXT: This morning, during a house meeting with you, Lisa, and Brian, you tetchily state that you have ended up tidying for Lisa for the last three weeks and refuse to do it again. Noting your irritation, Lisa volunteers to take the lead on cleaning the kitchen when she gets home from work this evening. **Brian seems satisfied with this plan.**

In the afternoon, while Lisa is still at work, you and Brian are standing in the kitchen doorway, gazing in. Brian turns to you and jokingly says,

UTTERANCE: ‘Can you clean the kitchen?’

If the reader remembers Brian’s perspective, namely that he ‘**seems satisfied with this plan**’ (i.e., for Lisa to do the cleaning), then the reader just needs to hold this contextual information in memory when interpreting the contrasting utterance (‘Can you clean the kitchen’) to understand that Brian is not seriously asking ‘you’ to act. In conversation, people commonly construct their utterances to be understood by their current conversational partner, a linguistic formula called audience design (Horton & Greig, 2005). In this example, the speaker (Brian) has played on the knowledge common to ‘you’ and himself for humorous effect; if he made this comment to anyone without this shared knowledge, the utterance would have a different meaning. Related to this idea, Horton and Greig (2005) suggest that conversation does not rely on mental representations; instead, smooth dialogue relies on conversational partners assessing common ground information and forming their utterances with this information in mind. Holding common ground knowledge in memory that is shared between characters may be particularly important to deciphering meaning in a text when other perspective information (e.g., facial expressions) is limited. This idea is supported by the findings in Study 3, which determined that working memory predicted performance in deployment. Therefore, detecting a speaker’s intentions to interpret the

meaning of a social pragmatic utterance in short narrative texts may be more reliant on language skills, working memory, and attention to become sensitive to speakers' perspectives than the ability to fully represent the given speaker's perspective.

This study examined a young adult demographic (18 to 40 years old). However, it is possible that age would have an effect on pragmatic language in older adults. In older age (over 65 years old), declines in the comprehension of sarcasm have been detected. Philips et al. (2015) found that older adults (between 65 and 86 years old) were significantly worse at interpreting sarcasm via videos and vignettes than younger (18-39) or middle-aged adults (40-64). Notably, middle-aged adults scored highest on indirect and direct forms of sarcastic language, indicating that the social experience that comes with age may aid in interpreting non-literal humour up until late middle age. Philips et al. (2015) study suggests that this decline in pragmatic ability is not due to working memory, as they only found weak correlations between working memory and written sarcasm comprehension. Although mentalising supported sarcasm comprehension in all age groups, there were no age effects in mentalising in this study. However, meta-analyses have shown a decline in mentalising in older adults (Henry et al., 2013). As mentalising has been shown to support pragmatic language, it may be that a decline in mentalising can explain a decline in pragmatic language comprehension in older age (Marocchini & Domaneschi, 2022; Trott & Bergen, 2020). Therefore, it may be possible to elicit the effects of mentalising on pragmatic inference when comparing younger and older adults.

In the current study, the results show that mentalising was not recruited to interpret the intended meaning of the utterance in the ironic comment condition. This finding contrasts with previous research in Italian (Marocchini & Domaneschi, 2022). The reasons for such contradictory findings remain unclear, and they may be due to cultural differences in pragmatic language. Marocchini and Domaneschi (2022) highlighted that ironic

comments were read at a significantly slower pace regionally and across the whole sentence in comparison with two control conditions. In addition, they found that higher scores on the mentalising task (Strange Stories) predicted faster reading times. The results of the present study are perplexing as the narratives in the ironic comment condition in the current experiment were closely based on those used in Marocchini and Domaneschi's (2022) study, which included vocabulary (e.g., 'teased', 'jokingly'; see example below) in the context that may have signalled that the forthcoming target utterance was ironic.

CONTEXT A: Giovanni has organised a party at his house. Francesco and Alessio discuss what to bring to the party.

CONTEXT B: It's a difficult choice because Giovanni does not drink any alcohol, and he is frequently teased by his friends because of it. Alessio jokingly says to Francesco:

IRONIC UTTERANCE: Can you bring some wine?

One reason the pattern of findings did not mirror those of Marocchini and Domaneschi (2022) may be due to language and cultural differences between English and Italian. Goddard (2018) notes that verbs such as 'joking' and 'teasing' are, in fact, more UK-English-specific than is commonly realised, which is indicative of the differences in the use of irony between people from the United Kingdom and Italian people. Currently, there is little research exploring the cultural differences between differences in the use of irony in the English and Italian languages. When searching the British and Italian press and online forums, Taylor (2016) suggests that the stereotype of the British using more irony in common language is false and that irony and sarcasm are used equally in both countries. However, it is unclear if there are qualitative differences in the way irony is used, and this would be an interesting avenue for future research.

Finally, the finding that ironic comments and conventionalised indirect requests are processed with comparable efficiency is intriguing, as the properties of each type of utterance are intrinsically different. Conventionalised indirect requests contain lexical cues that trigger the meaning of the utterance, whereas ironic comments do not. Furthermore, research is divided as to whether the processing of conventionalised indirect requests is context-dependent or context-independent (Breheny et al., 2013). For example, it is feasible that we can read the utterance ‘Can you pass the salt?’ alone without context and derive the indirect directive meaning (i.e., ‘Pass the salt’), but to derive an ironic interpretation of this utterance, context is necessary. However, if we assume that the intended meaning of the utterance in the ironic comment condition was derived before the sentence was read in full, the similarity in reading time patterns between the conditions may be attributed to the fact that in both conditions, the intended interpretation of the utterance was identified very rapidly, and therefore it was not possible to distinguish individual differences between them via the online methods. One difference between this study’s method and Marocchini and Domaneschi’s study is that here, key pieces of information regarding the speaker’s perspectives were highlighted in bold, which may have meant that the participants interpreted the intended meaning more quickly. To record differences in reading times between the two conditions, it may be necessary to use more sensitive instruments, such as eye-tracking.

6. General Discussion

This research aimed to broaden the understanding of individual differences in mentalising in adults in relation to reading comprehension. Conceptual development of mentalising skills (e.g., false belief understanding) in young children has been associated with reading comprehension both concurrently and longitudinally (Atkinson et al., 2017; Kim 2016, 2017). However, less is understood about when and how mentalising is recruited to support reading comprehension in skilled readers.

Chapter 1 reviewed existing research on the continuous development of mentalising after the age of five, which shows that the ability to use mentalising abilities flexibly is increasingly supported by other cognitive skills (e.g., language, memory). Furthermore, different cognitive skills may support mentalising in different contexts. In children, it has been suggested that mentalising supports understanding the mental states of characters in fictional texts and that the use of mentalising abilities during reading is supported by a number of non-mentalising skills (e.g., working memory, vocabulary, grammar; Dore et al., 2018; Kim, 2016, 2017). Therefore, when studying the effect of mentalising on reading comprehension in adults, it is important to (a) have a clear understanding of what different

advanced mentalising measure, and (b) how other cognitive skills contribute to the understanding of characters' mental states in text.

This thesis addressed four research questions: (1) What is the role of mentalising in reading comprehension in adults? (2) What subcomponents of mentalising do current mentalising tasks measure? (3) How are these subcomponents differentially related to reading comprehension in adults? (4) What mechanisms underpin the relationship between mentalising and reading comprehension in adults? To explore these four questions, four studies utilising a variety of methods were conducted. In Study 1, associations between one commonly used mentalising measure (Strange Stories) and reading comprehension were found. Study 2 used a systematised review to develop a taxonomy of advanced mentalising tasks. Studies 3 and 4 focused on pragmatic inference generation during reading (in the context of highly indirect requests and ironic utterances respectively) as potential mechanisms where mentalising might be recruited during reading. The results of these studies again identified Strange Stories as the mentalising task most closely associated with reading comprehension, specifically recruited during the sampling and maintenance of characters' mental states. However, given that the Strange Stories task was the only mentalising measure that was related to reading comprehension in these studies, it is unclear whether 'mentalising' per se is important, or whether shared pragmatic language comprehension demands explain the relationship. This discussion begins by summarising the key findings of each study before discussing their theoretical implications, identifying limitations and suggesting avenues for future research.

6.1 Summary of findings

6.1.1 Chapter 2: Investigating the Relationship Between Mentalising and Reading Comprehension in Adults: A Correlational Study (Study 1)

Study 1 investigated the association between mentalising skills and reading comprehension in adults, with the aim of testing whether mentalising is more strongly related to comprehension of literary texts than expository texts. This involved completing four mentalising tasks (visuospatial perspective-taking, attributions from non-verbal cues, interpreting social situations, and a task that draws on pragmatic language), two reading comprehension tasks based on either literary or expository texts, and two control tasks measuring vocabulary and inference generation.

Contrary to expectations, mentalising (Strange Stories) predicted unique variance in reading comprehension of expository texts but not literary texts after controlling for age, vocabulary, and inference generation. No test of fiction reading comprehension that is standardised on UK adults could be sourced; therefore, it is possible that psychometric weaknesses in the measure of fiction reading used partly explain this unexpected finding. There was also notably limited association between the four mentalising tasks used, highlighting the lack of coherence between the diverse measures used to assess mentalising beyond early childhood (Warnell & Redcay, 2019). Overall, the findings provide an initial indication that mentalising may be implicated in skilled reading comprehension.

6.1.2 Chapter 3: Properties of Mentalising Tasks Used to Measure the Relationship Between Mentalising and Language in Adults. A Systematic Review and Taxonomy of Mentalising Tasks

In light of the findings of Study 1, Study 2 aimed to explore the construct validity of currently used advanced mentalising tasks. A systematic review was conducted to (1) collate the mentalising tasks that had been used to study the relationship between mentalising and verbal skills, including reading comprehension, in adults since 2002, (2) identify which subcomponents of mentalising each task measures, and (3) examine the reported correlations between mentalising and language abilities. The results revealed that there is wide heterogeneity in the mentalising tasks used in research with adults. Of the 182 studies examined, 91 original mentalising tasks were found, yet only seven tasks accounted for 68% of uses (Reading the Mind in the Eyes, Strange Stories, Faux Pas Task, Awareness of Social Inference Test, Hinting Task, Frith-Happé Animations, Movie for the Assessment of Social Cognition). A taxonomy was created to systematically describe the properties of advanced mentalising measures. It is clear that these tasks assess different subcomponents of mentalising and draw on different non-mentalising skills. Despite this heterogeneity, the mentalising tasks were often used interchangeably between studies. This study indicates the need for greater precision in measurement of mentalising skills beyond early childhood.

In this review, correlations between mentalising and language measures were collated across studies. Only a small number of studies (41 studies out of 182, 22.5%) directly reported correlation coefficients. Weighted means effect sizes showed that there were significant positive correlations between four commonly used mentalising tasks (the Reading the Mind in the Eyes Test, the Strange Stories task, the Faux Pas Task, and the

Frith-Happé animations) and a range of language abilities (e.g., vocabulary, reading comprehension, pragmatics). For example, correlations between the Reading the Mind in the Eyes task and pragmatic language skills, such as figurative language, implicatures, and proverbs were identified (Bambini et al., 2021; Peterson & Miller, 2012). However, there were some confounding variables, which should be taken into account when considering the relationship between mentalising and language abilities. The participants' first language, country of origin and cultural norms varied across the studies. Language and cultural norms have been shown to affect the development of mentalising, which means that the associations between mentalising and language differ across countries (Henry et al., 2013). For example, making explicit mental state judgements about mental states may be more common in some cultures than others (Yong et al., 2022) (see Chapter 3, Section 3.5).

6.1.3 Chapter 4: When Do Comprehenders Mentalise for Pragmatic Inference? A Partial Replication Study

After establishing an association between mentalising and reading comprehension in adults in Studies 1 and 2, Study 3 aimed to investigate the mechanisms underlying this relationship in more detail. One way in which mentalising may aid reading comprehension is by supporting readers' understanding of the pragmatic components of text, such as indirect speech. To interpret the true meaning of indirect speech, the reader must make a pragmatic inference by combining the literal meaning of the utterance with information from contextual cues (e.g., mental state information). Mentalising may be recruited to interpret contextual cues about the speaker's mental states. To investigate this, Study 3 replicated and extended an experiment by Trott and Bergen (2020) exploring the relationship between mentalising and a form of written pragmatic language (indirect requests).

Trott and Bergen (2020) tested Apperly's (2018) component model of perspective taking, in which he suggests people recruit mentalising during three stages of pragmatic inference: when *sampling* the words in a speaker's utterance, *maintaining* it in memory, and finally when *deploying* this information to interpret the meaning of a pragmatic utterance. In theory, pragmatic language could be involved in some of these or all three stages. Trott and Bergen (2020) determined that mentalising, as measured using the Short Story Task, predicted performance on the pragmatic inferences during sampling and maintenance but not deployment, indicating that mentalising supports the interpretation of pragmatic language during reading.

This study included three notable extensions to Trott and Bergen's (2020) design. (1) The Strange Stories task was used to measure mentalising instead of the Short Story Task, which was found to be an unreliable measure of mentalising in Study 1. (2) A visual mentalising task, namely the Frith-Háppe Animations, was included to test whether the relationship between mentalising and language is specific to language-based mentalising tasks or whether it generalises to other types of nonlanguage-based mentalising tasks. (3) a working memory control measure was added, as previous research has shown that working memory may explain the association between mentalising and pragmatic language (Horton & Grieg, 2005; Long et al., 2018). The results of Study 3 indicated that the text-based (Strange Stories) but not the visuospatial mentalising measure (Animations) predicted performance on pragmatic inferences during sampling and maintenance, supporting the findings of Trott and Bergen (2020). These findings suggest that mentalising supports the building of a situational model of the text, in relation to the sampling and maintenance of characters' mental states. Post-hoc analyses revealed that working memory explained individual differences in deployment of characters' perspectives. This study provides novel

insights into the complex relationship between mentalising and pragmatic language comprehension during reading.

6.1.4 Chapter 5: When Do Comprehenders Mentalise for Pragmatic Inference? Experiment 2

Study 3 suggested that mentalising supports the first two stages of Apperly's (2018) component model of perspective-taking, sampling, and maintenance, but not the final stage of deployment. However, in the component model of perspective-taking, Apperly (2018) proposes that mentalising supports the deployment of mental state information for pragmatic inference—a finding supported by previous research (Trott & Bergen, 2020 – Experiment 2; Van Ackeren et al., 2016). Therefore, Study 4 used an alternative paradigm to study the relationship between mentalising and deployment for pragmatic inference. Specifically, this study examined the processing of ironic comments during reading in comparison to two control measures: conventionalised indirect requests and literal questions. The Silent Films Task was substituted for the Animations task as a second measure of mentalising, assessing whether the association between mentalising and aspects of reading comprehension extends beyond the text-based Strange Stories task. The Silent Films Task was designed as a companion to the Strange Stories, as it tests the same mentalising concepts (e.g., false beliefs, misunderstandings) but does not draw on reading comprehension. Contrary to expectations, the results of Study 4 did not show that ironic comments took longer to process than the control sentences, suggesting that metacognition (awareness of one's own thought processes) was not needed to process ironic comments. In addition, neither mentalising task predicted performance on this task. The results of Study 4 suggest that

mentalising does not support the deployment of pragmatic inference when reading ironic statements. However, it is also possible that the influence of mentalising was not detected due to methodological issues with the Pragmatic Language Task used in the study.

6.2 The role of mentalising in reading comprehension in adults

Mentalising aids individuals' understanding of their own and others' mental states (Premack & Woodruff, 1978). Dore et al.'s (2018) framework proposes that mentalising plays a key role in understanding the mental states of characters when reading fictional texts, because without the ability to envision others' perspectives, it would be difficult for readers to understand characters' behaviours in narrative texts. For example, in literary novels, the actions that occur in the narrative are often driven by the inner motivations of the characters (Bruner, 1986). Dore et al. (2018) further propose that mentalising may help the reader create coherent situational models of the text. According to this view, mentalising aids in the building of situational models during reading by enabling readers to understand the characters' perspectives. By creating clear representations of both the characters' mental states and goal-directed actions, the reader can make more accurate inferences about the character's actions and consequently gain a deeper understanding of the text.

6.2.1 Evidence of associations between mentalising and reading comprehension in adults

Study 1 was a validation study that first aimed to establish a relationship between different indices of mentalising and reading comprehension in adults, since previous studies focus on children (Atkinson et al., 2017; Boerma et al., 2017; Kim, 2016, 2017). To test Dore's framework, Study 1 also explored whether mentalising supports reading expository texts (factual text) and fictional texts (stories based on interpersonal relationships) differently. Due to the lack of research on adults, it was important to establish which aspects of mentalising may be related to comprehension of different types of text beyond childhood. Three different components of mentalising were targeted in Study 1, selected due to their prominence in research on reading abilities in children and adults: visuospatial perspective-taking, attributions from non-verbal cues, and mental state reasoning (Boerma et al., 2017; Kidd & Castano, 2013; Zwaan & Radvansky, 1998). Four tasks were used: the Director's Task (visuospatial perspective-taking), the Reading the Mind in the Eyes Test (attributions from non-verbal cues), and two text-based measures, the Strange Stories (mental state reasoning, pragmatic language) and the Short Story Task (mental state reasoning, interpreting social situations). To control for overlapping task demands, two language control measures—vocabulary and inference generation—were included. Finally, two reading comprehension tasks—assessed literary (Short Story Task, reading comprehension subscale) and expository text comprehension (ART-2) respectively. Because no standardised test of fiction text comprehension in adults was available, the reading comprehension subsection of the Short Story Task was used.

Contrary to predictions and the findings of childhood research, the results of the regression analysis in Study 1 suggested that mentalising (Strange Stories) predicted a unique variance in reading comprehension of expository texts (ART-2) but not literary texts (Atkinson et al., 2017; Boerma et al., 2017; Kim, 2016, 2017). Because mentalising allows individuals to reason about the thoughts and feelings of others, it was expected that it would

support reading fiction texts, which tend to explore human emotions and social situations.

Contrastingly, expository texts are more factual and less likely to recruit mentalising.

However, this predicted pattern was not observed.

One possible explanation is that mentalising does indeed support reading comprehension of fictional texts more than expository texts, but this was obscured due to measurement issues. The Short Story Task reading comprehension subscale showed poor internal reliability (Cronbach's $\alpha = .31$). It is a short (five-item), non-standardised test, and factor analysis showed that the items did not reliably load onto a single factor. A clearer relationship between mentalising and narrative reading comprehension may well be observed using a more sensitive and reliable measure. Additionally, it may be that the topic of the text used in the Short Story task was not of interest to all the participants, which may have affected engagement, particularly in the context of online administration. Research shows that when individuals are allowed to choose texts, they are interested in they show more motivation to engage (Alzubi & Nazim, 2024). Moreover, the measure of expository text comprehension (ART-2) included some information about historical characters, their actions and motivations. For example, Story 1 describes a group of researchers who go to Africa to investigate wildlife. The ART-2 inference-based items draw upon the motivations of the researchers, and mentalising may be recruited to answer these questions. Item 3 asks, 'Why did the scientists go to Africa?' and item 9 asks, 'Why do you think they want to make a film?' Mentalising may support inference-making when answering these questions due to their social nature (Dore et al., 2018).

Alternatively, it may be that mentalising indirectly supports the development of expository reading comprehension across developmental stages. Mentalising may support reading comprehension of fictional texts during childhood development, which leads to better reading comprehension of all texts. Indeed, a mutually reinforcing relationship may

exist between mentalising and reading comprehension. Mentalising may help children understand characters' mental states in fictional texts at a young age (Atkinson et al., 2017; Boerma et al., 2017; Kim, 2016; 2017); this improved understanding of texts may in turn encourage them to read more, which improves their oral language skills (e.g., vocabulary, narrative comprehension) and, consequently, their reading comprehension of all text types (Cunningham & Stanovich, 1991; Jackson, 2022; Mar et al., 2006; Mol & Bus, 2011).. Thus, it may be that mentalising continues to support reading comprehension of fictional texts across the lifespan, which in turn improves reading comprehension of all text types.

On the other hand, the Strange Stories task draws on other cognitive and language skills beyond mentalising, and these may account for the correlation between scores on the Strange Stories Task and ART-2 in Study 1. To account for this, Study 1 controlled for two key skills that support reading comprehension (vocabulary and inference generation), and Strange Stories was still shown to predict unique variance in reading comprehension ability (Cain et al., 2001; Kim, 2017). Notably, previous research that has shown relationships between mentalising and reading has done so using verbal mentalising tasks (e.g., false belief tasks; Atkinson et al., 2017; Boerma et al., 2017; Ebert, 2020; Kim, 2016, 2017), pointing to the possibility of a special relationship between reasoning about mental states *verbally* and reading comprehension.

However, there are three notable skills associated with reading comprehension that were not controlled for in Study 1. Firstly, the lexical decision task was used to measure vocabulary breadth (quantity of vocabulary knowledge) but not vocabulary depth (quality of vocabulary knowledge). However, both vocabulary depth and breadth have been associated with reading comprehension (Ouellette et al., 2006). Second, both mentalising and reading comprehension draw on working memory, which may explain the association between mentalising and reading comprehension (Kim, 2017; Wardlow et al., 2013). Third,

individual differences in grammar processing at a young age have been shown to predict performance on the Strange Stories in middle childhood (Ebert, 2020). Therefore, the association between the Strange Stories and reading comprehension, could be based on linguistic (vocabulary depth, grammar) and cognitive (working memory) skills associated with reading rather than mentalising.

Overall, the findings from Study 1 suggest that an association exists between mentalising and reading comprehension in adults but leaves several questions unanswered. More specifically, the results indicate that one component of mentalising measured using the Strange Stories (language-mediated mental state reasoning) explained unique variance in reading comprehension. However, the hypothesis that mentalising plays a special role in the comprehension of fictional texts could not be meaningfully assessed, due to measurement issues. In addition, it remains unclear whether the observed relationship between the Strange Stories and ART-2 is due to the mentalising or shared language and/or working memory demands between the tasks.

6.3 What do mentalising tasks actually measure?

Currently there is little research on what advanced mentalising tasks measure, although there is a growing body of work showing a lack of coherence between different mentalising measures beyond the age of five, meaning that performance on one mentalising task does not predict performance on another (Białecka-Pikul et al., 2021; Warnell & Redcay, 2019). The findings reported in this thesis support this claim. In Study 1, small to no correlations were observed between adult performance on the Director's Task, Reading the Mind in the Eyes, the Strange Stories and Short Stories tasks. Similarly, Study 3 revealed little correlation between two widely used mentalising tasks, the Strange Stories

and the Frith-Happé Animations, possibly because they measured different mentalising concepts (see Section 6.4.2). In addition, there were stronger correlations between the control conditions of the Strange Stories and the Frith-Happé Animations, indicating that non-mentalising properties of the tasks (e.g., vocabulary, visuospatial skills) explain more variance than the mentalising components. On the other hand, Study 4 indicated that there was a moderate correlation between the Strange Stories and the Silent Films Task ($r = .35^{**}$). However, the Silent Films Task was designed as a companion for the Strange Stories, which measures the same mentalising concepts but does not rely upon reading skills. Therefore, a stronger correlation would be expected. Nevertheless, the Strange Stories and the Silent Films Task appear to cohere more than other mentalising tasks. Although the Silent Films task shows concurrent validity with the Strange Stories in young children and adolescents (7–13 years old), there is little research showing the task's concurrent validity in adults (Devine & Hughes, 2013; 2016).

Given this uncertainty about measurement of advanced mentalising, the systematic review (Study 2) had two primary aims: (1) to examine how mentalising has been operationalised when exploring the relationship between mentalising and language in adults, and (2) to examine the properties of the different mentalising tasks and form a taxonomy of the mentalising concepts that each task measured. The most notable finding of this review was that, while many measures exist, a small selection of are commonly used to operationalise mentalising in adults: the Reading the Mind in the Eyes, Strange Stories, Faux Pas Task, Awareness of Social Inference Test, Hinting Task, Frith-Happé Animations, and the Movie for the Assessment of Social Cognition. Despite the evolution of mentalising theory, researchers are still using tasks that were designed in the 1990s and early 2000s to measure mentalising as a unidimensional construct. There is very little agreement in the studies included in the review about which components of mentalising each task measures.

This lack of consensus is best exemplified by descriptions in the literature of the Reading the Mind in the Eyes Test, which has been variously defined as measuring : attribution of complex emotional and cognitive states (Meinhardt Injac et al., 2020), mental state inference (Turkstra et al., 2008), advanced affective theory of mind (TOM) (Demers & Koven, 2015), emotional recognition (Ahmed & Miller., 2011), visual TOM (Spencer et al., 2020), and advanced meta-representational skills (Zegarra Valdivia et al., 2019). Notably, few researchers explicitly explained their reasoning for defining the Reading the Mind in the Eyes Test as they did. The task was originally designed to test the ability to attribute mental states on the basis of facial cues (an aspect of mentalising). However, research has demonstrated that the task may be more accurately described as measuring emotional perception. Recent metanalytic research by Kittel et al. (2022) indicates that the Reading the Mind in the Eyes Test correlates more strongly with measures of emotional perception than theory of mind. To uncover how aspects of mentalising work in combination with other non-mentalising skills, such as language, it is imperative to systematically identify and catalogue the various subcomponents of mentalising and specify which subcomponent of mentalising existing tasks measure. This was the purpose of building the taxonomy presented in Chapter 3.

Another notable finding from the systematic review was that researchers often use a single task to measure mentalising. Eighty out of the 182 included studies (44%) used a single mentalising measure, and in a third of these, the Reading the Mind in the Eyes Test was the only mentalising measure used. There are two flaws to this approach. (1) It assumes mentalising is a unitary construct, yet as shown by the taxonomy, the task only measures a single component of mentalising. The fact that the authors commonly do not specify that the Reading the Mind in the Eyes Test measures a single subcomponent of mentalising means that the construct validity of the task is ambiguous and open to interpretation. This is a

particular problem in mentalising research as, up until the last two decades, mentalising was widely considered a unidimensional rather than multidimensional construct (Frith & Frith, 2006; Warnell & Redcay, 2019). (2) It ignores the fact that, as shown by the taxonomy, all mentalising tasks draw on other, non-mentalising cognitive skills. For example, the Reading the Mind in the Eyes Test draws upon vocabulary (Rigby et al., 2016) and visual processes, such as face perception (Deska & Hugenberg, 2017). It is, therefore important that multiple measures of advanced mentalising are incorporated in study designs, so that it can be determined which specific mentalising subcomponents and non-mentalising cognitive skills are associated with the phenomenon of interest.

6.4 Does mentalising support reading comprehension through pragmatic language?

In Study 3, it was proposed that one way in which mentalising may support readers' understanding of the characters' mental states is by supporting the interpretation of pragmatic language in text. A relationship between mentalising and pragmatic language has been long established in research. However, most of this research focuses on the relationship between mentalising and oral language (Grice, 1975). Hulme and Snowling (2014) have indicated that pragmatics, along with semantics and grammar, contribute to the language comprehension component of the Simple View of Reading. The systematic review in Study 2 identified studies showing an association between mentalising and written pragmatic

language comprehension (Ahmed & Miller 2013; Bambini et al., 2021). It is, therefore, plausible that mentalising supports the oral language component of the Simple View of Reading via pragmatics.

The primary focus of Studies 3 and 4 was to investigate whether mentalising has a special role in processing pragmatic language in written text, which in turn supports reading comprehension. Specifically, mentalising may improve readers' understanding of characters' spoken dialogue, which involves understanding the intentions behind speakers' utterances as well as the literal meaning. When interpreting spoken dialogue, it has been suggested that listeners build a mental representation of the speakers' perspectives to make more accurate inferences about the meaning of speakers' utterances (Apperly, 2018). Apperly's component model of perspective-taking, posits that mentalising is recruited over three stages when inferring the meaning of spoken dialogue: firstly, when *sampling* the words in a speaker's utterance; secondly, when *maintaining* it in memory; and finally, when *deploying* this information to interpret the meaning of pragmatic language. Studies 3 and 4 were designed to assess the proposition that mentalising improves the reader's understanding of a character's dialogue by enabling more accurate judgements about the character's intentions. This would help readers to monitor and represent the mental states of characters more accurately during narrative processing and make more accurate pragmatic inferences.

6.4.1 Apperly's (2018) component model of perspective-taking

Interpreting a pragmatic utterance is a complex process as the words in spoken dialogue alone are often not sufficient to express the speaker's intended meaning. To do this, one must become sensitive to another's perspective, such as what the speaker sees, hears, knows, or believes, in order to place the speaker's utterances in context. For example, if a

friend asks, ‘How’s your mum doing?’, there are several possible interpretations of this question. Perhaps they are just being polite; however, if you knew they knew your mother had recently been diagnosed with cancer, this same question could be interpreted as asking how she was dealing with the disease. To make a pragmatic inference about the meaning of an utterance, one becomes *sensitive* to the speaker’s perspective by attending to relevant information and memorising it (Horton & Grieg, 2005). However, if there is no shared common ground information to refer to (e.g., you are unsure whether your friend knows about your mother’s illness when they ask about her), one may need to consciously *represent* what a person knows or does not know and make a pragmatic inference about the meaning of their utterances.

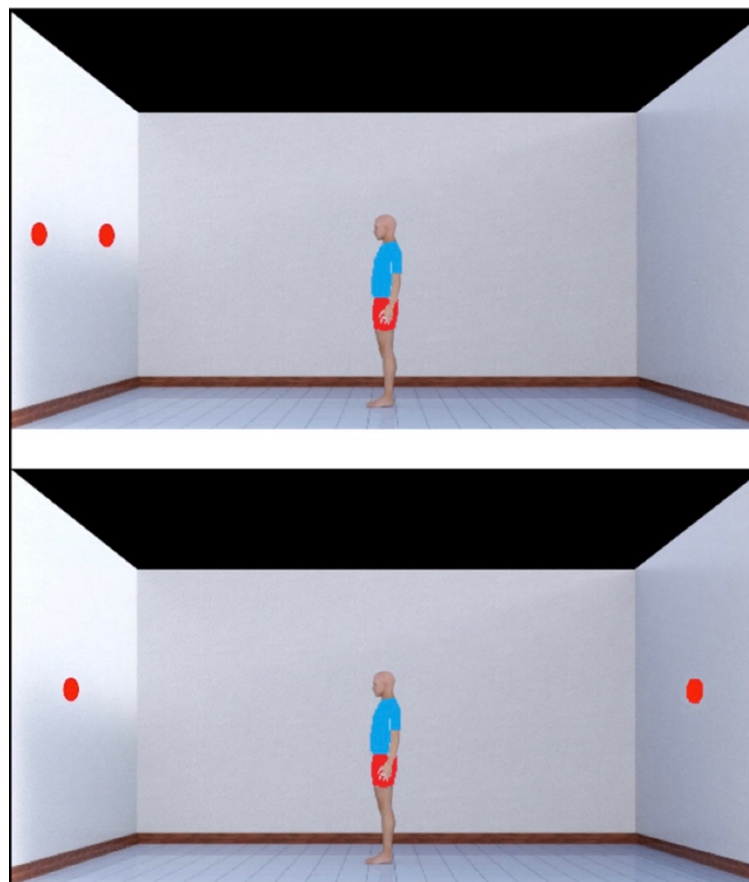
In many cases, it may be possible to coordinate these stages and track a co-communicator’s perspective using only attention and memory resources (Horton & Gerrig, 2005; Pickering & Garrod, 2004). However, it is also possible that mentalising may be needed to support the representation of a co-communicator’s perspectives during each stage (Keysar et al., 2000; Keysar et al., 2003). Apperly’s (2018) component model of perspective-taking offers a structured approach to understanding how individuals infer others’ mental states while making pragmatic inferences. Yet, so far, there is only a small amount of research that supports this theory. Apperly refers to his own research on using a referential communication task, the Director’s Task, as evidence that perspective-taking occurs in stages (sampling, maintenance, deployment). In the Director’s task the participant must sample the visual perspective of the director, then hold this information in memory, and finally deploy this information to infer which object the director intends the participant to move. However, the Director’s task relies on representing the director’s visual perspective and the task may be more reliant on visuospatial perspective taking skills, rather than taking the other’s mental state perspective (e.g., thoughts, emotions) (Cole & Millet,

2019; Rubio-Fernández, 2017). More evidence is needed to support the theory that readers use this same staged process to infer the meaning of characters' pragmatic language.

Additionally, although Apperly states that others' perspectives may be processed automatically or using conscious effort, it is unclear when each process is engaged. Critiques of Apperly's model suggest that the amount that people spontaneously take the perspectives of others during the sampling stage is overestimated. Cole and Millett (2019) have criticised studies (e.g., Samson et al., 2010), which employed the "dot perspective task" (See Figure 6.1) to show evidence of spontaneous computing of others' visual perspectives.

Figure 6. 1

The Dot Perspective Task (Samson et al., 2010)



Note. In this paradigm, participants must say how many dots are presented. Reaction times are generally shorter when the agent sees the same number of dots as the participant (upper panel), than when the participant sees a different number to the agent.

Cole and Millet (2019) suggest these experiments lack rigorous controls, and results may be attributed to the use of domain-general skills (e.g., working memory and attention) rather than evidence of perspective-taking. This view is supported by Rubio-Fernández (2017), who suggests that performance on the Director's Task can be explained via selective attention processes rather than showing evidence of individuals consciously taking the

director's perspective. In summary, Apperly's component model provides a framework for understanding the mechanisms involved in taking other people's perspectives while making pragmatic inferences. However, more empirical research is needed to understand when automatic and conscious processes occur during this process.

6.4.2 Does mentalising support pragmatic inference?

In Studies 3 and 4, Apperly's component model of perspective-taking was used as a theoretical framework to investigate how characters' pragmatic language, specifically indirect requests and ironic statements, is processed during reading. Study 3 replicated Trott and Bergen's (2020) results, which indicate that Apperly's component model of perspective taking can be applied to written narrative texts. More specifically, mentalising supports the sampling of mental state information from the text and maintaining this information in memory, but deploying this information to make a pragmatic inference about the meaning of an indirect request relies on other skills. Post-hoc tests suggest that working memory supports the deployment stage.

Trott and Bergen's (2020) study was one of the first to present evidence of readers adopting a character's perspective during each of Apperly's component stages for pragmatic inference in a written fictional narrative. This study is important as it sheds light on the possible mechanisms behind the role of mentalising in pragmatic language in text. However, the mentalising measure used in Trott and Bergen's (2020) study was the Short Story Task, which was demonstrated to be an unreliable measure of mentalising in Study 1 of this thesis. In addition, the Short Story Task is text-based, and it is possible that its verbal properties are account for Trott and Bergen's findings. Moreover, the only control measure used by Trott and Bergen was the reading comprehension subsection of the Short Story Task, which was

also shown to be unreliable in Study 1. Therefore, to re-test Trott and Bergen's results, Study 3 was a replication and extension study that substituted two widely used tasks of advanced mentalising, the Strange Stories (text-based measure) and the Frith-Happé Animations (visual mentalising task). In addition, Trott and Bergen's study was extended here by adding working memory as a control variable; as Apperly indicates, it is possible to coordinate perspective with working memory and attention alone without the need for representation of the perspectives (Apperly, 2018). The results of Study 3 replicated Trott and Bergen's by showing that (1) altering the speaker's knowledge states systematically changed the rate at which the speaker's perspectives were sampled, maintained, and deployed; (2) participants sampled and maintained speaker's perspectives at a higher rate than they deployed this information for pragmatic inference; and (3) mentalising predicted sampling and maintenance but not deployment when using the Strange Stories as a predictor.

As the Strange Stories, but not the Frith-Happé Animations, predicted performance during sampling and maintenance, it was unclear whether this relationship was due to the mentalising or language properties of the Strange Stories. The Strange Stories Physical Stories task was used to control for reading comprehension, and it therefore seems unlikely that shared reading demands explain this association. While it is noteworthy that the Strange Stories Physical is not a standardised reading comprehension task, it utilises the same format and coding as the Strange Stories and thus should be a suitable control for reading comprehension (White et al., 2009). However, it is possible that pragmatic language elements of the Strange Stories task (rather than 'mentalising' per se) explain its association with the Pragmatic Language Task (Bosco et al., 2018a).

As the key skill in understanding pragmatic utterances is identifying the speaker's intentions, distinguishing between pragmatic language and mentalising is methodologically

challenging. Thus, to understand the relationship between mentalising and pragmatics in the Strange Stories and Trott and Bergen's indirect request task, it is necessary to explore which concepts each task draws upon. In Trott and Bergen's indirect request task, to successfully interpret the speaker's intentions, the participant must identify what the speaker knows or does not know. For instance, do they know about an obstacle (e.g., a car heater is broken) to fulfilling a request (e.g., to turn on the car's heating)? Being able to make judgements about another person's states of knowledge or ignorance is a mentalising concept that is usually acquired in young children before the age of five years old (Apperly, 2012; Wellman & Lui, 2004). On the other hand, Strange Stories measures individual differences in people's ability to use a range of mentalising concepts (beliefs, false beliefs, desires) flexibly to understand the reasons why people say something or act a certain way. Using mental state concepts in this way may more accurately reflect how mental state reasoning is conducted in real life. Apperly (2012) argues that mental state constructs are not usually processed in isolation. Instead, an individual must use a coherent set of interrelated mental state concepts (e.g., knowledge, desires, beliefs) to understand other people's behaviour. Therefore, based on the results of this study, it is possible that individuals who are more skilled in using a range of mental states flexibly, as tested in the Strange Stories, are better at identifying a speaker's knowledge states when making pragmatic inferences about the meaning of indirect requests. Alternatively, it could be argued that both tests measure the participant's ability to judge the intended meaning of language, meaning that both tasks may be measuring similar constructs (i.e., pragmatic language).

As the Frith-Happé Animations (visual mentalising task) did not predict performance on the indirect request task, this may suggest that the association found between the Strange Stories (text-based mentalising task) and the indirect request task was due to shared verbal properties. However, drawing on the taxonomy presented in Chapter 3, it is apparent that the

Frith-Happé Animations and the Strange Stories measure different mentalising subcomponents. It is therefore impossible to conclude from the results of this study whether the mentalising or language properties (e.g., pragmatic language comprehension) of the Strange Stories are responsible for the association between the Strange Stories and the indirect request task. Examining the results further, there were stronger correlations between the mentalising measures and their control measures (e.g., Strange Stories Physical and Frith-Happé Goal-Directed animations) than between the mentalising measures themselves. This could indicate that the non-mentalising aspects of the Strange Stories (e.g., vocabulary, inferencing, grammar) and the Frith-Happé Animations (visuospatial processing, expressive language) may explain more variance between the tasks than any shared mentalising properties. In support of this, research suggests that there is little coherence between narrative-based tasks, such as the Strange Stories, and animacy perception tasks, like the Frith-Happé Animations, as narrative tasks draw on social understanding and language use in social settings, whereas animacy perception tasks draw on visuospatial processing skills (Rasmussen & Jiang, 2019; Scholl & Goa, 2013).

6.4.3 The role of working memory in deployment

Post-hoc tests suggest that deploying perspective information to make a pragmatic inference about the meaning of an indirect request is supported by working memory. In Apperly's (2018) component model of perspective-taking, it is proposed that adopting a speaker's perspective during sampling, maintenance, and deployment may only rely on their ability to become sensitive to the speaker's perspectives through attention and memory

rather than having to fully represent their perspectives. This study demonstrated that mentalising supported sampling and maintenance but not deployment (which drew on working memory). In light of these findings, it is reasonable to suggest that, once a situational model has been formed, readers may only need to access this information via memory to infer the meaning of pragmatic utterances, such as indirect requests. However, this may only be true in the context of interpreting short, written vignettes, as in the current experimental paradigm. When reading long texts (e.g., literary novels), it may be necessary to fully represent the speaker's utterances in the context of larger themes (e.g., thematic inferences; Graesser et al., 1994).

In conclusion, the results of Study 3 replicated the main findings of Trott and Bergen (2020). The results suggest that people adopt the speaker's perspective during sampling, maintenance, and deployment when making a pragmatic inference during reading. Moreover, altering the knowledge states of the speaker has a larger effect on the rate at which people sample and maintain perspective information than when deploying this information to make a pragmatic inference. Mentalising (as measured by a more reliable task than in Trott and Bergen's study: Strange Stories) supports the construction of a situational model of the text, when sampling information about a character's perspectives and maintaining this information in memory, but not when deploying this perspective information to interpret the meaning of a pragmatic utterance. However, the Frith-Happé Animations test did not predict performance during sampling and maintenance; it is possible that the association between the Strange Stories and indirect request task was due to the shared pragmatic language demands rather than mentalising per se.

6.5 Does mentalising have a role in the deployment of pragmatic inference?

In light of the findings from Study 3, Study 4 aimed to further investigate the involvement of mentalising in deployment using a different experimental paradigm. This study investigated whether mentalising supports deployment in an alternative form of pragmatic inference (i.e., comprehension of ironic comments). Participants performed a self-paced, region-by-region reading task where they were presented with sentences starting with the words ‘Can you’ presented in an ironic, literal question, or conventionalised indirect request form. The target sentences were the same in each condition, and the correct interpretation of the utterances was based on cues in the context, such as information about speakers’ perspectives. Reading times were calculated locally across each of the three regions and globally across the whole sentence. Participants’ mentalising ability was measured via the Strange Stories (text-based mentalising) and the Silent Films Task (visual mentalising task).

The reading time results of this study did not fit the predictions and suggest that meta-representation is not needed for deployment for pragmatic inference in the ironic comment condition (Marocchini & Domaneschi, 2022). At a local level, there was no evidence of longer reading times in Region 3 in the ironic comment condition, which would indicate wrap-up effects (Haberlandt & Graesser, 1985; Stine-Morrow et al., 2010). Moreover, longer reading times were noted in Region 1, indicating that the reader encoded the meaning of the ironic comments when reading this initial part of the sentence. However, ironic comments do not contain linguistic cues at the start of the sentence to indicate the utterance’s meaning. It is therefore likely that the readers were over-primed to expect the ironic comment before processing the ironic utterance in full (Gildea & Glucksberg, 1983).

In support of this, global reading times were found to be significantly shorter in the ironic condition than in the literal condition, further indicating that increased processing was not necessary to interpret the ironic comments. Taken together, the local and global reading time results suggest that the ironic comments were not meta-represented in this study.

Apperly's (2018) component model of perspective-taking posits that priming, memory, and attention processes may be enough to track and become sensitive to communicators' perspectives. In the ironic comment condition of the pragmatic language paradigm, readers may become sensitive to the speaker's perspective by noting words that prime the reader to a particular interpretation. If this is the case, it would not be necessary for the reader to represent the speaker's perspective. For example, below is an abbreviated excerpt from a narrative in the ironic comment condition.

CONTEXT: This morning, during a house meeting with you, Lisa, and Brian, you tetchily state that you have ended up tidying for Lisa for the last three weeks and refuse to do it again. Noting your irritation, Lisa volunteers to take the lead on cleaning the kitchen when she gets home from work this evening. **Brian seems satisfied with this plan.**

In the afternoon, while Lisa is still at work, you and Brian are standing in the kitchen doorway, gazing in. Brian turns to you and jokingly says,

UTTERANCE: 'Can you clean the kitchen?'

In this excerpt, three people in a house are trying to decide who should clean it. 'You refuse' to clean it, so Lisa says she will do it in the evening. If we assume that to interpret the meaning of an ironic comment, one must meta-represent the speaker's perspective with the linguistic meaning of their utterances, then one must note the juxtaposition between Brian's

perspective (he **seems satisfied with the plan** ... of Lisa cleaning tonight) and Brian's words in the utterance ('Can you clean the kitchen?') to interpret the speaker's meaning. However, it is also possible that the reader is primed to take the ironic meaning of the words by 'jokingly says', and in this case it is no longer necessary to take Brian's perspective. Indeed, the reading times suggest that longer processing times, indicating greater processing, took place in region 1 of the ironic utterance, which would indicate it was unnecessary to read the whole sentence to grasp the speaker's intended ironic meaning, suggesting that meta-representation did not take place. Therefore, it appears logical to conclude that the readers were over-primed to understand the ironic meaning.

Apperly's (2018) component model of perspective-taking proposes that the speaker's perspectives need to be represented during deployment. However, Apperly's model was built to explain spoken interactions. Thus, it is possible that when reading shorter passage-length texts, deployment for pragmatic inference is processed differently. In shorter text passages, cues to the speakers' intentions are restricted; in contrast to spoken dialogue, readers do not have access to many of the paralinguistic signals (e.g., intonation, facial expressions, and eye contact) that may help the reader become sensitive to the speaker's perspectives. The reader is therefore more reliant on linguistic cues in the text to interpret the meaning of dialogue. Vygotsky (1962) notes that written language is generally aimed at an imaginary or absent person; in contrast, spoken language is usually tailored to a specific interlocutor. Oral language is dynamic and led by the speakers' current thoughts and feelings, so it is possible that listeners need to be attentive to the mental states of their interlocutor to understand them (Schallert et al., 1977). Due to the different properties of oral and written language, it may be that the deployment of pragmatic inference during spoken language is more reliant on mentalising than when reading. The results of Study 4 suggest that, during reading, words that prime the reader toward the meaning of an utterance

(e.g., ‘jokingly says’) are sufficient for successful comprehension; meta-representation is not necessary. It is possible that when these priming words are unavailable, readers may need to represent the perspectives of the speaker during deployment. Future research should therefore consider removing such cues (priming words) to explore the role of mentalising when reading passage-length texts.

The findings from Study 4 contrast with those of Marocchini and Domaneschi (2022), who used similar methods to demonstrate that mentalising supported the deployment of pragmatic inference when interpreting ironic comments. They found ironic comments were processed more slowly than the control conditions and indicated that meta-representation of the ironic comments during deployment took place. In addition, the Strange Stories predicted performance at this stage. The narratives used in Study 4 were very similar to those used by Marocchini, and they also included vocabulary that may have caused over-priming (e.g., ‘teased’ and ‘jokingly’; see example below).

CONTEXT A: Giovanni has organised a party at his house. Francesco and Alessio discuss what to bring to the party.

CONTEXT B: It's a difficult choice because Giovanni does not drink any alcohol, and he is frequently teased by his friends because of it. Alessio jokingly says to Francesco:

IRONIC UTTERANCE: ‘Can you bring some wine?’

The reasons for these different results may be due to language and cultural differences. Research translating spoken dialogue from English to Italian suggests that humour is often lost in translation and that literal translations of wordplay may hinder comprehension of humorous intentions (Antonini, 2005; Antonini et al., 2003). Therefore, the language in Marocchini’s translation may be exaggerated to ensure that an English audience understands

the intended meaning. Currently, there is little research exploring the use of irony in English compared to Italian. Therefore, future research should explore the intricacies of irony and whether these differ across languages and cultures.

6.5.1 Conclusion: Mentalising and pragmatic language

Studies 3 and 4 support the application of Apperly's (2018) component theory of perspective-taking to explain the role of mentalising in pragmatic inference in written text. The findings from Studies 3 and 4 suggest that mentalising supports the first two stages of pragmatic inference, namely sampling mental state information from the text and maintaining it in memory, but not the third phase of deploying this information for pragmatic inference. This supports Trott and Bergen's (2020) findings from Experiment 1 of their study (replicated in Study 3) and indicates that mentalising supports the construction of a situational model of the text and holding it in memory. On the other hand, mentalising is not drawn upon when using this information to make a pragmatic inference about the meaning of an utterance; this step relies on other skills, such as working memory. The findings of Studies 3 and 4 are useful as they give insights into putative mechanisms underpinning the relationship between mentalising and reading comprehension. However, the findings of Study 4 (that mentalising does not support deployment) need to be considered with caution in light of the methodological problems in the study (over-priming).

6.6 Limitations and considerations for future research

In the studies reported in this thesis, multiple measures of mentalising were included in order to attempt to specify which specific subcomponents are involved in reading comprehension. However, only a select number of mentalising tasks were used across the three empirical studies (Chapters 2, 4, and 5). Of the tasks used, only Strange Stories was predicted reading comprehension and pragmatic language across all studies. There are two possible interpretations of these results. (1) The subcomponents of mentalising captured by the Strange Stories (e.g., understanding double bluffs, white lies, misunderstanding) have a particular relationship with reading comprehension and pragmatic language. In support of this, past research has associated understanding concepts captured by the Strange Stories (such as characters' higher-order false beliefs) with reading comprehension in children and adults (Atkinson et al.; Kim, 2016; 2017; Rickett's et al., 2013; Trott & Bergen, 2020). (2) The non-mentalising skills drawn upon by the Strange Stories (vocabulary, inferencing, pragmatic language) are responsible for the relationship between the Strange Stories and the reading comprehension and pragmatic tasks (Bosco et al., 2017).

The use of Strange Stories as a mentalising predictor is potentially more problematic when examining the relationship between mentalising and pragmatic language. Pragmatic language examines dialogue in shared communication, either in spoken language or written text (Dewart & Summers, 1995). The Strange Stories examines participants' ability to reason about the mental states of characters in social situations and explain why they acted a certain way or said something that was not literally true, which draws on pragmatic language. As Strange Stories draws on pragmatic language, there is debate as to whether it is a suitable measure to investigate the relationship between mentalising and pragmatic

language comprehension during reading due to the shared characteristics of the tasks (Bosco et al., 2018a). In support of this, studies that have only used the Strange Stories' false belief items (deemed as 'pure' mentalising concepts) to avoid the confounding pragmatic elements of the Strange Stories have found weaker correlations with pragmatic language than studies that use the full version (Ahmed & Miller, 2011; Gavilán & García-Albea, 2011; White et al., 2009). The pragmatic language elements of the Strange Stories could therefore explain stronger correlations. Although these different results could also be explained by language and task differences, it may be better to use mentalising tasks that do not tap pragmatic language to investigate the relationship between the constructs. The tasks selected in the current studies (Frith-Happé Animations; Silent Films) did not relate to performance on reading comprehension tasks.

A limitation of Study 2 (systematic review) is that few papers reported correlational results between mentalising tasks and language abilities. Studies 3 and 4 were designed on the basis of the associations between mentalising and pragmatic language tasks in Study 2, but evidence from the review was scant and inconclusive. It is possible that mentalising supports some forms of pragmatic language more than others. In Study 2, a total of seven correlations were reported between mentalising and pragmatic language across four different papers. Two studies reported correlations between the Reading the Mind in the Eyes Test and figurative language, implicatures and a general pragmatic language task (which examines figurative language and humour). Also, two studies reported correlations between the Strange Stories and metaphors, irony and proverbs. These results suggested that there were moderate significant correlations between the Reading the Mind in the Eyes Test and pragmatic language (weighted mean = .43, $p < .001$) and small correlations between the Strange Stories and pragmatic language (weighted mean = .24, $p < .05$). However, even in this small pool of correlational results there was substantial variation in the size of the

correlations. For example, Gavilán and García-Albea (2011) found no relationship between the Strange Stories and Proverbs ($r = .01$). In contrast, Ahmed and Miller (2011) found moderate correlations ($r = .40, p < .05$) between the same two tasks. This variation is likely due to methodological and/or cultural/language differences. For example, Gavilán and García-Albea (2011) only used an abbreviated version of the Strange Stories, which contained only three items related to false beliefs. As this task only contains three items, the task may have low reliability (Field, 2018). Tests with too few items are prone to higher measurement error as unpredicted participant variability can disproportionately affect the overall score (DeVellis, 2017). Therefore, the reason Gavilán and García-Albea (2011) found no correlation between the Strange Stories and proverbs may be that they used a test with too few items. Despite this, more research is needed to establish correlations between mentalising and pragmatic language in the English language.

Pragmatic language encompasses a range of different forms (e.g., humour, indirect requests, implicatures) and more research is needed to explore how mentalising supports each form of pragmatic language. In Studies 3 and 4, it was found that mentalising predicted some stages of processing indirect requests but that mentalising did not support the processing of ironic comments. It is likely there are methodical issues (e.g., over-priming) in Study 4 that may have been responsible for the null results. However, it is also possible that mentalising supports the interpretation of indirect requests more than ironic comments, which may also explain why mentalising predicted pragmatic language in Study 3. Currently, there is little research exploring the relationship between mentalising and pragmatic language in typically developing children and adults and beset my measurement issues. Existing studies tend to focus is on the relationship between mentalising and specific forms of pragmatic language (e.g., indirect requests, humour) (Marocchini & Domaneschi, 2022; Trott & Bergen, 2020). In future, a review comparing the relationship between

mentalising and different forms of pragmatic language would be useful to improve precision of measurement and inform future research.

Study 4 suffered from methodological issues with the paradigm used to elicit the effects of mentalising during deployment in the ironic comment condition; this means that there is still some uncertainty surrounding these results. In this study, participants read narratives and were then meant to encode the intended meaning of an utterance by interpreting the literal meaning of the utterance in combination with the characters' perspectives (e.g., knowledge states, desires, intentions). However, when considering reading times locally in the ironic comment condition, it appeared that participants were interpreting the intended meaning of the utterance after reading the region of the sentence ('Can you'), which indicated they were not reading the sentence in full before understanding the intended (ironic) meaning of the utterance, likely because they were over-primed by the context.

As such, the methods used in this study may not have been sensitive enough to elicit the effects of mentalising due to the methodological issues mentioned above. Adjusting the methods used in this study may enable researchers to elicit the effects of mentalising during the deployment of pragmatic inference.

6.6.1 Future directions

Considering the strengths and limitations of this thesis and existing work, several future directions would help to elucidate the role of mentalising in reading comprehension and pragmatic language. First, despite research stressing that literacy in adults is an immediate

concern in the United Kingdom, research into the mechanisms behind reading comprehension in adults is lacking (Dugdale & Clark, 2008). This is evident in the lack of standardised reading comprehension tests available for adults, particularly tests evaluating the comprehension of fictional texts. Additionally, this thesis provides evidence of the potential role of mentalising in reading comprehension and pragmatic language. However, there are few mentalising measures available that are specifically designed for adults. The studies in this thesis revealed that measures such as the Strange Stories and the Silent Films Tasks were able to elicit individual differences in mentalising in adults, yet these tasks were designed for children in middle childhood and early adolescence. Hence, mentalising measures need to be developed that can reliably assess the comprehension and use of more sophisticated mental state concepts, which are appropriate for capturing individual differences in adult samples. The Short Story Task has attempted to do this, but the format and coding system render this task unreliable. Specifically, a task that tests adults' ability to flexibly use the mental state concepts (false beliefs, misunderstandings, double bluff) drawn upon by the Strange Stories in a more sophisticated manner would be useful.

6.6.2 Conclusion

The findings reported in this thesis add to a growing body of work supporting the proposition that mentalising has a role in reading comprehension. Furthermore, this is one of the first studies to produce evidence of this relationship in adults. Taken together, the findings indicate that mentalising is recruited during the interpretation of pragmatic language, for example, in the reported dialogue between characters, which supports the

construction of a situational model of the text. Apperly's (2018) component model of perspective-taking can be adapted as a framework to explain how people make pragmatic inferences during reading. Mentalising may support the sampling of explicit mental state information from the text and the maintenance of this information in memory to facilitate the building of coherent situational models. The information from these situational models can then be deployed to infer the meaning of pragmatic sentences. Further empirical work and more sensitive measures are needed to support these conclusions. Nonetheless, the findings have implications for models of adult reading comprehension.

7. Appendices

Appendix A - Pragmatic Language Task materials (Study 3)

Pragmatic Language Task

Practice Trials

Literal Question condition

Narrative

After you finish university, you start working in a large office building. The work is pretty mundane data entry stuff but the people in the office are nice, and you get on well with your boss, Richard. In order to get from floor to floor most people take the lift but you like to take the stairs as it keeps you fit.

While running down the stairs you slip and twist your ankle. Luckily you are on your floor and manage to hobble across to your desk in pain. **Your boss Richard sees you hobbling in pain** as he is walking towards you with a file.

Richard is concerned for you. As a manager he has to be very careful about health and safety around the office. He tells you he was going to ask you to take a file upstairs. You reach for the file, but he withdraws it from your reach, looks at your ankle, then **seriously** asks,

Experimental question

Can you take this file upstairs

Irony condition

Narrative

After you finish university, you start working in a large office building. The work is pretty mundane data entry stuff but the people in the office are nice, and you get on well with your boss, Richard. In order to get from floor to floor most people take the lift but you like to take the stairs as it keeps you fit.

While running down the stairs you slip and twist your ankle. Although you are on your floor you manage to take a few steps then have to sit down before reaching your desk. **Your boss Richard sees you sat on the floor in pain and clutching your ankle** as he is walking towards you with a file.

He tells you not to move and waves over to ask one of the other members of staff to help. You are pretty embarrassed about it all and apologise to him profusely. As you are waiting for help you ask him why he was coming over to you. He says he needed to get a file to finance. You both chuckle, and he then **jokingly** says,

Experimental question

Can you take this file upstairs

Indirect Request condition

Narrative

After you finish university, you start working in a large office building. The work is pretty mundane data entry stuff but the people in the office are nice, and you get on well with your boss, Richard. In order to get from floor to floor most people take the lift but you like to take the stairs as it keeps you fit.

While running down the stairs you slip and twist your ankle. Luckily you are on your floor and manage to hobble across to desk and sit down to recover.

You see your boss Richard, **who is not aware of your accident**, walking towards you with a file. He tells you he needs to get an important file up to finance ASAP. He holds out the file for you to take and says,

Experimental question

Can you take this file upstairs

Pragmatic Language Task

Experimental Trials

Story 1

Literal Question condition

Narrative

You've been renting a house with your two roommates, Lisa and Brian, for three months now. Things are going well, though both you and Lisa are not very responsible about cleaning; consequently, the house has a tendency to get dirty quickly - unwashed dishes in the sink, cluttered counters, the whole works.

Even worse, you and Lisa are directly responsible for the latest mess, which is the product of a recent dinner party. Brian's rather obsessive about neatness, so he's pretty anxious about the mess.

This morning, during a house meeting with you, Lisa, and Brian, you state that **you are too busy today to clean the house** as you are **working from home**. So, Lisa volunteered to take the lead on cleaning the kitchen when she gets home from work this evening. Brian seems satisfied with this plan.

In the afternoon, while **Lisa is still at the office**, you **take a break from work**. While heading to the kitchen you see Brian standing in the kitchen doorway, gazing in. Brian, who **knows you're busy and that you may not be able to spare any time** hopefully asks,

Experimental question

Can you clean the kitchen

Irony condition

Narrative

You've been renting a house with your two roommates, Lisa and Brian, for three months now. Things are going well, though both you and Lisa are not very responsible about cleaning; consequently, the house has a tendency to get dirty quickly - unwashed dishes in the sink, cluttered counters, the whole works.

Even worse, you and Lisa are directly responsible for the latest mess, which is the product of a recent dinner party. Brian's rather obsessive about neatness, so he's pretty anxious about the mess.

This morning, during a house meeting with you, Lisa, and Brian, you tetchily state that you have **ended up tidying for Lisa** for the last three weeks and **refuse to do it again**. Noting your **irritation** Lisa volunteered to take the lead on cleaning the kitchen when she gets home from work this evening. **Brian seems satisfied with this plan.**

In the afternoon, while Lisa is still at work, you and Brian are standing in the kitchen doorway, gazing in. Brian turns to you and jokingly says,

Experimental question

Can you clean the kitchen

Indirect Request condition

Narrative

You've been renting a house with your two roommates, Lisa and Brian, for three months now. Things are going well, though both you and Lisa are not very responsible about cleaning; consequently, the house has a tendency to get dirty quickly - unwashed dishes in the sink, cluttered counters, the whole works.

Even worse, you and Lisa are directly responsible for the latest mess, which is the product of a recent dinner party. Brian's rather obsessive about neatness, so he's pretty anxious about the mess. This morning, Lisa volunteered to take the lead on cleaning the kitchen when she gets home from work this evening. **Brian was still sleeping, so he's unaware of this plan.**

In the afternoon, while Lisa is still at the office, you and Brian are standing in the kitchen doorway, gazing in. Brian turns to you and says,

Experimental question

Can you clean the kitchen

Story 2

Literal Question condition

Narrative

You and your friend are exhausted and sweaty from playing tennis. After playing, you invite him back to your apartment to hang out.

When you return to your apartment, you both see a notice taped to your apartment door: unfortunately, the water in your apartment building will be shut off for the next 24 hours. Your friend seems disconcerted by the news. You tell him not to worry, they **probably** haven't **turned everything off just yet** and you'll have to **test it**.

Once inside the kitchen, you head to the kitchen sink. Your friend turns to you **inquiringly** and asks,

Experimental question

Can you pour me some water

Irony condition

Narrative

You and your friend are exhausted and sweaty from playing tennis. After playing, you invite him back to your apartment to hang out.

When you return to your apartment, you both see a notice taped to your apartment door:
unfortunately, the water in your apartment building will be shut off for the next 24 hours. Your friend seems disconcerted by the news.

Once inside the kitchen, your friend, still drenched in sweat, **turns the knob on the kitchen tap but the water does not flow**, he turns to you and **jokingly** says,

Experimental question

Can you pour me some water

Indirect Request condition

Narrative

You and your friend are exhausted and sweaty from playing tennis. After playing, you invite him back to your apartment to hang out.

When you return to your apartment, **you go to your room**, and privately check your email. You see that your landlord has sent you a message: unfortunately, the water in your apartment building will be shut off for the next 24 hours. **Your friend, of course, doesn't know about this yet.**

Once inside the kitchen, you head to the kitchen sink. Your friend turns to you and says,

Experimental question

Can you pour me some water

Story 3

Literal Question condition

Narrative

After several weeks of looking for work, you finally find a job at a local cafe. You quickly learn the basics: taking orders from customers, how to prepare different drinks, and even which items are the most popular.

Customers especially love their famous chocolate chip muffins, but the baker stops making new ones by noon, so these are often in short supply.

One afternoon, a regular elderly customer enters the store. Just as he approaches the counter, you both hear a shout from the back room: "Hey, we're **running low** on chocolate chips for the muffins, so don't offer any more muffins as we may **not be able to make any more** for the customers." **The customer looks agitated when he hears this.**

He **inquiringly** asks,

Experimental question

Can you get me a muffin

Irony condition

Narrative

After several weeks of looking for work, you finally find a job at a local cafe. You quickly learn the basics: taking orders from customers, how to prepare different drinks, and even which items are the most popular.

Customers especially love their famous chocolate chip muffins, but the baker stops making new ones by noon, so these are often in short supply.

One afternoon, a regular elderly customer enters the store. Just as he approaches the counter, you both hear a shout from the back room: “Hey, we’re out of muffins!” The customer looks **disappointed when he hears this**. He looks up at the menu and his eyes alight on the pastry section. You give him a wide smile and ask what he would like today, you say he can have anything except for the muffin.

He looks at you and with a smile says,

Experimental question

Can you get me a muffin

Indirect Request condition

Narrative

After several weeks of looking for work, you finally find a job at a local cafe. You quickly learn the basics: taking orders from customers, how to prepare different drinks, and even which items are the most popular.

Customers especially love their famous chocolate chip muffins, but the baker stops making new ones by noon, so these are often in short supply.

One afternoon, the cafe runs out especially early, and you hear a shout from the back room: “Hey, we’re out of muffins!”

Soon after, a regular elderly customer enters the store, who is **unaware that you've just run out of the most popular item, muffins**. You exchange pleasantries. He looks up at the menu and his eyes alight on the pastry section. He says,

Experimental question

Can you get me a muffin

Story 4

Literal Question condition

Narrative

Your partner, Kate, has been very busy recently with extra-long days at work, coming home late in the evening. As a consequence, laundry has started to pile up, which Kate doesn't really have the time to do.

One evening, after Kate has returned home feeling exhausted, you decide it's time to make a dent in some of the laundry. But when you try to turn on the washing machine, you realize that it is broken. While trying to fix it, you **slice your hand with a screwdriver**. Later, after bandaging your hand up and with **Kate watching**, you make an appointment for someone to fix the machine, and **she seems satisfied**.

Later that day, you notice Kate glancing with concern at the ominously large laundry hamper. You tell her there is no way to get her shirts cleaned by tomorrow, other than hand-washing them. Kate looks at **your bandaged hand**, and then **seriously** asks,

Experimental question

Can you wash the clothes

Irony Condition

Narrative

Your partner, Kate, has been very busy recently with extra long days at work, coming home late in the evening. As a consequence, laundry has started to pile up, which Kate doesn't really have the time to do.

One evening, after Kate has returned home, you decide it's time to make a dent in some of the laundry. But when you try to turn on the washing machine, you realize that it is broken. With Kate watching, you make an appointment for someone to come fix the machine.

Kate knows that you made an appointment to fix the broken washing machine and seems satisfied. The laundry hamper is growing ominously large at this point. Seeing it, **Kate picks up a small bar of soap from the sink and tosses it to you**, then **laughs** and says,

Experimental question

Can you wash the clothes

Indirect Request condition

Narrative

Your partner, Kate, has been very busy recently with extra long days at work, coming home late in the evening. As a consequence, laundry has started to pile up, which Kate doesn't really have the time to do.

One day, while Kate is at work, you decide it's time to make a dent in some of the laundry. But when you try to turn on the washing machine, you realize that it is broken. You make an appointment for someone to come fix the machine.

Soon after, Kate returns home exhausted. **Kate does not know that the washing machine is broken.**

The laundry hamper is growing ominously large at this point. Seeing it, Kate says,

Experimental question

Can you wash the clothes Story 5

Literal Question condition

Narrative

You and your friend Paul are moving into a new apartment. There's a ton of stuff to unload: pots and pans, furniture, miscellaneous belongings, and more. Most of it is light, but some of the furniture is quite heavy. For the heavy stuff, you agree to help each other out.

At one point, while Paul is helping you carry your mattress into your bedroom, you suddenly feel a jolt of sharp pain in your back. You inadvertently gasp and clutch your back. Once you recover, you decide you should take it easy the rest of the day and not pick up any more heavy furniture.

Knowing you are hurt, Paul finishes pushing the mattress into the room.

The two of you go back outside. Paul notices you walking awkwardly clutching your lower back as you go. Outside you see Paul's massive oak desk sitting in the driveway. **Paul looks at the desk doubtfully**, and then **looks at you with concern** and asks,

Experimental question

Can you help me carry this

Irony condition

Narrative

You and your friend Paul are moving into a new apartment. There's a ton of stuff to unload: pots and pans, furniture, miscellaneous belongings, and more. Most of it is light, but some of the furniture is quite heavy. For the heavy stuff, you agree to help each other out.

At one point, while Paul is helping you carry your mattress into your bedroom, you suddenly feel a jolt of sharp pain in your back. You inadvertently gasp and clutch your back. You decide you should take it easy the rest of the day and not pick up any more heavy furniture. **Knowing you are hurt,** Paul finishes pushing the mattress into the room.

The two of you go back outside. **Paul notices you walking awkwardly clutching your lower back as you go.** Outside you see Paul's massive oak desk sitting in the driveway. Paul walks to the desk and slaps the surface, with a **large grin** he says,

Experimental question

Can you help me carry this

Indirect Request condition

Narrative

You and your friend Paul are moving into a new apartment. There's a ton of stuff to unload: pots and pans, furniture, miscellaneous belongings, and more. Most of it is light, but some of the furniture is quite heavy. For the heavy stuff, you agree to help each other out.

At one point, while Paul is still outside and you are carrying your mattress into your room, you suddenly feel a jolt of sharp pain in your back. You inadvertently gasp and clutch your back. Once

you recover, you decide you should take it easy the rest of the day and not pick up any more heavy furniture. **Paul, however, didn't see what happened, so he doesn't know that you are hurt.**

When you go back outside, you see that Paul is still out front, looking at his massive oak desk. He says,

Experimental question

Can you help me carry this

Story 6

Literal Question condition

Narratives

Tonight is new year's eve, so you, and your two flatmates, Aili and Olivia, are deciding how you should celebrate. You all had tickets to a club, but the night was cancelled at the last moment due to the DJ pulling out. So, Aili suggests throwing a party in the flat. The three of you start planning a big shop to get some food and drinks for that night.

Later that day, the three of you go to the supermarket. You take your car as it has a lot of boot space. At the supermarket, you get food and lots of alcohol. You and Olivia take the full shopping trolley across to your car as Aili goes to the kiosk to buy some cigarettes.

When you get to the car, you open the boot and find that it is full of clothes that you and Olivia were going to take to a charity shop. You and Olivia stare at the open boot. Then, Aili comes up behind you with a couple of bags in each hand. As he approaches, **Aili sees that the boot is full of** clothes and realises **you may not be able to** fit anything else in it. With a **concerned look** he asks you,

Experimental question

Can you fit this in

Irony condition

Narratives

Tonight is new year's eve, so you, and your two flatmates, Aili and Olivia, are deciding how you should celebrate. You all had tickets to a club, but the night was cancelled at the last moment due to the DJ pulling out. So Aili suggests throwing a party in the flat. The three of you start planning a big shop to get some food and drinks for that night.

Later that day, the three of you go to the supermarket. You take your car as it has a lot of boot space. At the supermarket, you get food and lots of alcohol. You and Olivia take the full shopping trolley across to your car as Aili goes to the kiosk to buy some cigarettes.

When you get to the car you open the boot and find that it is full of clothes that you and Olivia were going to take to a charity shop. You both stare at the open boot. Then, Aili comes up behind you with a couple of bags in each hand. As he approaches, **Aili sees that the boot is full of clothes, and that it will be impossible to fit the shopping in.** At this point, you all start laughing, with a large grin Aili says,

Experimental question

Can you fit this in

Indirect Request condition

Narrative

Tonight is new year's eve, so you, and your two flatmates, Aili and Olivia, are deciding how you should celebrate. You all had tickets to a club, but the night was cancelled at the last moment due to the DJ pulling out. So Aili suggests throwing a party in the flat. The three of you start planning a big shop to get some food and drinks for that night.

Later that day, the three of you go to the supermarket. You take your car as it has a lot of boot space.

At the supermarket, you get food and lots of alcohol. You and Olivia take the full shopping trolley across to your car as Aili goes to the kiosk to buy some cigarettes.

When you get to the car you open the boot and find that it is full of clothes. Then, Aili, **who can not see that the boot is full**, comes up behind you with a couple of bags in each hand, and says,

Experimental question

Can you fit this in

Story 7

Literal Question condition

Narrative

You and your friend, Karen, have lived together for two years. You get on really well as roommates: one of the reasons for this has been that you always split all the bills equally and each pay your share, so you never argue about money.

This month though a large chunk of your wage has suddenly disappeared, as you have had a student loan repayment come out of your bank account that you had forgotten about. You realize you cannot afford to pay the rent.

Today is the day that one of you normally pays the landlord. As you are looking at your balance on your phone in the living room, Karen comes into the living room. She tells you she has transferred £450 into your account for her half of the rent.

You tell her about your money problems. Karen gives you a hug and tells you that sucks and says she'll make you both a cup of tea. As she gets up, **you see a concerned look suddenly flash across her face**, she turns to you seriously and asks,

Experimental question

Can you pay the rent

Irony condition

Narrative

You and your friend, Karen, have lived together for two years. You get on really well as roommates: one of the reasons for this has been that you always split all the bills equally and each pay your share, so you never argue about money.

This month though a large chunk of your wage has suddenly disappeared, as you have had a student loan repayment come out of your bank account that you had forgotten about. You realize you cannot afford to pay the rent.

Today is the day that one of you normally pays the landlord. As you are looking at your balance on your phone in the living room, Karen comes into the living room. She tells you she has transferred **£450 into your account for her half of the rent.**

You tell her about your money problems and then you ask Karen if there is any way she could lend you the money. Karen says that's fine, then pulls five pounds out of her pocket, passes it to you and smiling says,

Experimental question

Can you pay the rent

Indirect Request condition

Narratives

You and your friend, Karen, have lived together for two years. You get on really well as roommates: one of the reasons for this has been that you always split all the bills equally and each pay your share, so you never argue about money.

This month though a large chunk of your wage has suddenly disappeared, as you have had a student loan repayment come out of your bank account that you had forgotten about. You realize you cannot afford to pay the rent.

Today is the day that one of you normally pays the landlord. As you are looking at your balance on your phone in the living room, Karen, **who is unaware of your money troubles**, comes into the living room. She tells you she has transferred £450 into your account for her half of the rent, and then says,

Experimental question

Can you pay the rent

Story 8

Literal Question condition

Narratives

Over the summer you go back to visit your mum on her birthday. Her favourite thing to do is to go out into the country for long walks. So when you get back you plan a day trip to the dales in Yorkshire as she loves the views there. As you get your things together, she tells you to bring the new camera she bought you at Christmas because she wants to get some panoramic pictures while you're out there.

While you are walking, your mum tells you to make sure you get some photos of her at Malham Cove as that's where the best views are. When you get to a nice spot at Malham Cove your mum puts her rucksack on the floor and looks for a nice spot for a picture. Unfortunately, as you are looking at the camera you realize you did not charge it up and the battery is **almost** dead.

You tell your mum this and show her the camera. She attempts to turn it on, but the screen stays black. After repeatedly pressing the buttons on the camera with no success, she huffs realising that she is **not able to turn it on**, and then doubtfully asks,

Experimental question

Can you take a photograph

Irony condition

Narrative

Over the summer you go back to visit your mum on her birthday. Her favourite thing to do is to go out into the country for long walks. So when you get back you plan a day trip to the dales in Yorkshire as she loves the views there. As you get your things together, she tells you to bring the new camera she bought you at Christmas because she wants to get some panoramic pictures while you're out there.

While you are walking your mum tells you to make sure you get some photos of her at Malham Cove as that's where the best views are. Unfortunately, as you are looking at the camera you realize you did not charge it up and **the battery is dead**. You **tell your mum this** who looks disappointed but tells you to forget about it.

When you get to a nice spot at Malham Cove you and your mum admire the view. **She then chuckles to herself** and says,

Experimental question

Can you take a photograph

Indirect Request condition

Narrative

Over the summer you go back to visit your mum on her birthday. Her favourite thing to do is to go out into the country for long walks. So, when you get back you plan a day trip to the dales in Yorkshire as she loves the views there. As you get your things together, she tells you to bring the new camera she bought you at Christmas as she wants to get some panoramic pictures while you're out there.

As you are walking, your mum tells you to make sure you get some photos of her at Malham Cove as that's where the best views are. Unfortunately, as **you are looking at the camera you realize** you did not charge it up and the battery is dead. Your mum **doesn't know** the camera is dead.

When you get to a nice spot at Malham Cove your mum puts her rucksack on the floor, and then poses in front of the landscape and says,

Experimental question

Can you take a photograph Story 9

Literal Question condition

Narrative

Your friend Maria is in the final year of her English degree. She is a very good student but finds it impossible to manage her workload, so she always ends up finishing her work last minute and handing it in minutes before the deadline. Today is the deadline for her to hand in her dissertation, which needs to be printed off and submitted in paper format.

Maria messages you an hour before the deadline that she is a couple of minutes away and needs you to print off two copies of her dissertation. This is the first time she has mentioned this. Unfortunately, **your printer has been out of order for several days.**

You hear the doorbell ring and open the door. Immediately, Maria storms past you and goes into the living room where the printer is and starts trying to turn it on. Nervously, **you tell her that it's been acting up and you think it might be broken** and so you **may not be able to print anything.** Maria looks at you with a terrified expression, then turns and stares intensely at the printer, with a quaking voice she asks,

Experimental question

Can you print my paper

Irony condition

Narratives

Your friend Maria is in the final year of her English degree. She is a very good student but finds it impossible to manage her workload, so she always ends up finishing her work last minute and handing it in minutes before the deadline. Today is the deadline for her to hand in her dissertation, which needs to be printed off and submitted in paper format.

Maria messages you **three hours before** the deadline that she is a couple of minutes away and needs you to print off two copies of her dissertation. This is the first time she has mentioned this.

Unfortunately, your printer has been playing up for a while, and regrettably **in a fit of rage you hit it, smashing the glass and breaking the printer.**

You hear the doorbell ring and open the door. Maria heads into the living room where the printer is and starts trying to turn it on. Nervously, **you tell her that it is broken.** Maria looks at the broken glass and says she can see that. When you tell her what happened she bursts out laughing. Relieved she is not annoyed, you ask her if there's anything you can do to help, and still chuckling she says,

Experimental question

Can you print my paper

Indirect Question condition

Narrative

Your friend Maria is in the final year of her English degree. She is a very good student but finds it impossible to manage her workload, so she always ends up finishing her work last minute and handing it in minutes before the deadline. Today, is the deadline for her to hand in her dissertation, which needs to be printed off and submitted in paper format.

Maria messages you an hour before the deadline that she is a couple of minutes away and needs you to print off two copies of her dissertation. This is the first time she has mentioned this. Unfortunately, **your printer has been out of order for several days**. However, Maria **does not know this**.

Later that day, you hear the doorbell ring. When you open the door Maria storms past you to the living room where the printer is kept. You follow her into the living room and as you enter, she hands you her dissertation and says,

Experimental question

Can you print my paper

Story 10

Literal Question Condition

Narrative

Today you head over to your friend Oliver's flat to help him move out, as he has found a new flat. You haven't been to his flat before and you find he lives in a high storey building. You give him a ring and he tells you to head up to the tenth floor.

Unfortunately, you have claustrophobia and so you can't go into enclosed spaces like lifts, or you will start to panic. You take the stairs up, which is very tiring. When you finally get to his floor, you see that Oliver already has a lot of stuff ready to take down.

Oliver **knows about your claustrophobia** and so you may **not be able to take the lift** so asks,

Experimental question

Can you take the lift

Irony condition

Narrative

Today you head over to your friend Oliver's flat to help him move out, as he has found a new flat. You haven't been to his flat before and you find he lives in a high storey building. You give him a ring and he tells you to head up to the tenth floor.

Unfortunately, you have claustrophobia and so you can't go into enclosed spaces like lifts, or you will start to panic. You take the stairs up. When you finally get to his floor, you see that Oliver already has a lot of stuff ready to take down.

Oliver **knows about your claustrophobia** and has always been **very understanding about it as he realizes it's a serious condition**. He sees you breathing heavily from the walk up, and jokingly he

Experimental question

Can you take the lift

Indirect Request condition

Narratives

Today you head over to your friend Oliver's flat to help him move out, as he has found a new flat. You haven't been to his flat before and you find he lives in a high storey building. You give him a ring and he tells you to head up to the tenth floor.

Unfortunately, you have claustrophobia and so you can't go into enclosed spaces like lifts, or you will start to panic. You take the stairs up, which is very tiring. When you finally get to his floor you see that Oliver already has a lot of stuff ready to take down.

As you leave the flat, Oliver, **who does not know you have claustrophobia** says,

Experimental question

Can you take the lift

Story 11

Literal Question condition

Narratives

You and Kamala have been friends at university, and you have always had a crush on her. She is very intelligent and has always won awards during her mathematics degree. Due to this, you have always gone out of your way to help her.

In her final year, she needs to analyze her results using a computer software programme she is unfamiliar with. To try and help, you say you will do it and get it back to her by the deadline the next day. Kamala is very relieved by this.

Unfortunately, when you see the computer software package you realise you are way out of your depth with the material. When Kamala comes around the next day to get the work, she finds you frantically working away at it, you tell her you will email the results to her shortly. She looks at the **minimal amount of work you have completed so far** and realises **you may not be able to** finish it. Sounding concerned she asks,

Experimental question

Can you send me the results

Irony Condition

Narrative

You and Kamala have been friends at university, and you have always had a crush on her. She is very intelligent and has always won awards during her mathematics degree. Due to this, you have always gone out of your way to help her.

In her final year, she needs to analyze her results using a computer software programme she is unfamiliar with. To try and help, you say you will do it and get it back to her by the deadline the next day. Kamala is very relieved by this.

Unfortunately, when you see the computer software package you realise you are way out of your depth with the material.

Luckily, when Kamala comes round the next day, she tells you that she had a second look at it and managed to figure it out herself, she then sees a printout of the **minimal amount of work you have completed**. Embarrassed you tell her it was a bit harder than you expected, you both laugh. Picking up the piece of paper, she says,

Experimental question

Can you send me the results

Indirect Request condition

Narrative

You and Kamala have been friends at university, and you have always had a crush on her. She is very intelligent and has always won awards during her mathematics degree. Due to this, you have always gone out of your way to help her.

In her final year, she needs to analyze her results using a computer software programme she is unfamiliar with. To try and help, you say you will do it and get it back to her by the deadline the next day. Kamala is very relieved by this.

Unfortunately, when you see the computer software package you realise you are way out of your depth with the material. Despite, staying up all night you barely make a dent in the work.

The following day, you meet Kamala, who is **unaware you did not do the work**, and so presumes you have finished the analysis. Upon meeting her she says,

Story 12

Literal Question condition

Narrative

You have just started a new job as a writer at an online fashion magazine. Your parents are very pleased that you got this new job. As your computer is very old and slow, your parents have bought you a new laptop. Unfortunately, you spill coffee over it one morning and now it will not switch on.

Today your dad, dropped by to visit unexpectedly. So, you invite him in, and you both have a coffee in your living room. As you are sitting together his office rings. His colleagues want him to send across a spreadsheet with some financial information. He hates using computers and usually gets his secretary to do this kind of work for him. He takes out a pad that has a lot of notes he has made.

Your dad is **unaware the laptop he bought you is broken**, but needs to dictate some notes for you to type into a spreadsheet. However, he knows you are very bad at making spreadsheets and **may not be able to** use the computer to do this. He tells you he needs you to put some numbers into a spreadsheet, then cautiously asks,

Experimental question

Can you use your laptop

Irony condition

Narratives

You have just started a new job as a writer at an online fashion magazine. Your parents are very pleased that you got this new job. As your computer is very old and slow, your parents have bought you a new laptop. Unfortunately, you split coffee over it one morning and now it will not switch on.

Your dad opens the computer and attempts to turn it on. **Realising it's broken, your dad looks at you with a blank stare.** You tell him about the accident. Seeing how upset you are, your dad **tells you not to worry as it is covered by insurance.**

Relieved, you ask him what he was trying to do. He says he needed to make a spreadsheet, then he starts comically prodding at the keys on the dead laptop and laughs. He then jokingly says,

Experimental question

Can you use your laptop

Indirect Request condition

Narrative

You have just started a new job as a writer at an online fashion magazine. Your parents are very pleased that you got this new job. As your computer is very old and slow, your parents have bought you a new laptop. Unfortunately, you split coffee over it one morning and now it will not switch on.

Today your dad, dropped by to visit unexpectedly. So, you invite him in, and you both have a coffee in your living room. As you are sitting together his office rings. His colleagues want him to send across some financial information. Your dad hates using computers but knows you are good with them and so you **often write up his notes for him when he requests it.** He takes out a pad with a lot of notes he has made.

You dad **doesn't know your computer is broken**. He tells you he needs to dictate some notes for you to type up, and says,

Experimental question

Can you use your laptop

8. References

- Abell, F., Happé, F., & Frith, U. (2000). Do triangles play tricks? Attribution of mental states to animated shapes in normal and abnormal development. *Cognitive Development*, 15(1), 1-16. [https://doi.org/10.1016/s0885-2014\(00\)00014-9](https://doi.org/10.1016/s0885-2014(00)00014-9)
- Abu-Akel, A., Wood, S., Hansen, P., & Apperly, I. (2015). Perspective-taking abilities in the balance between autism tendencies and psychosis proneness. *Proceedings of the Royal Society London B*, 282, 1808. <https://doi.org/10.1098/rspb.2015.0563>
- Adams, A. G., Henry, J. D., Molenberghs, P., Robinson, G. A., Nott, Z., & von Hippel, W. (2019). The relationship between social cognitive difficulties in the acute stages of stroke and later functional outcomes. *Social Neuroscience*, 15(2), 158–169. <https://doi.org/10.1080/17470919.2019.1668845>
- Ahmed, F. S., & Stephen Miller, L. (2011). Executive function mechanisms of mentalising. *Journal of Autism and Developmental Disorders*, 41(5), 667–678. <https://doi.org/10.1007/s10803-010-1087-7>
- Ahmed, F. S., & Miller, L. S. (2013). Relationship between mentalising and functional independence is mediated by executive function. *Psychology and Aging*, 28(2), 293–303. <https://doi.org/10.1037/a0031365>
- Anderson, R. C., & Pearson, P. D. (1984). A schema-theoretic view of basic processes in reading comprehension. In P. D. Pearson, R. Barr, M. Kamil, & P. Mosenthal (Eds.), *Handbook of reading research* (pp. 255–291). New York: Longman.

- Andreasen, N. C., Calage, C. A., & O'Leary, D. S. (2008). Mentalising and schizophrenia: a positron emission tomography study of medication-free patients. *Schizophrenia bulletin*, 34(4), 708-719. <https://doi.org/10.1093/schbul/sbn034>
- Antonini, R. (2005). The perception of subtitled humor in Italy. *Humor*, 18(2).
<https://doi.org/10.1515/humr.2005.18.2.209>
- Antonini, R., Bucaria, C., & Senzani, A. (2001). It's a priest thing, you wouldn't understand: Father Ted goes to Italy. *Fuentes*, 82.
- Appel, M., & Mara, M. (2013). The persuasive influence of a fictional character's trustworthiness. *Journal of Communication*, 63(5), 912–932.
<https://doi.org/10.1111/jcom.12053>
- Apperly, I. A. (2012). What is “mentalising”? Concepts, cognitive processes and individual differences. *Quarterly Journal of Experimental Psychology*, 65(5), 825-839.
<https://doi.org/10.1080/17470218.2012.6760>
- Apperly, I. (2018). Mindreading and Psycholinguistic Approaches to Perspective Taking: Establishing Common Ground. *Topics in Cognitive Science*, 10(1), 133–139.
<https://doi.org/10.1111/tops.12308>
- Apperly, I. A. (2021). Cognitive basis of mindreading in middle childhood and adolescence. *Theory of mind in middle childhood and adolescence: Integrating multiple perspectives*, 37-54. <https://doi.org/10.4324/9780429326899-4>
- Arcara, G., & Bambini, V. (2016). A test for the assessment of pragmatic abilities and cognitive substrates (APACS): Normative data and psychometric properties. *Frontiers in Psychology*, 7, 70. <https://doi.org/10.3389/fpsyg.2016.00070>
- Artuso, C., & Belacchi, C. (2021). Semantic memory and reading comprehension: the relationship through adulthood and aging. *Aging Clinical and Experimental Research*, 1-11. <https://doi.org/10.1007/s40520-020-01771-1>

- Astington, J.W. & Jenkins, J.M. (1995). *Language and mentalising: a theoretical review and a longitudinal study*. Paper presented at the biennial meeting of the Society for Research in Child Development, Indianapolis, IN. <https://doi.org/10.1037/0012-1649.35.5.1311>
- Astington, J. W., & Jenkins, J. M. (1999). A longitudinal study of the relation between language and theory-of-mind development. *Developmental psychology*, 35(5), 1311. <https://doi.org/10.1037/0012-1649.35.5.1311>
- Astington, J. W., Pelletier, J., & Homer, B. (2002). Theory of mind and epistemological development: the relation between children's second-order false-belief understanding and their ability to reason about evidence. *New Ideas in Psychology*, 20(2-3), 131–144. [https://doi.org/10.1016/s0732-118x\(02\)00005-3](https://doi.org/10.1016/s0732-118x(02)00005-3)
- Atkinson, L., Slade, L., Powell, D., & Levy, J. P. (2017). Mentalising in emerging reading comprehension: A longitudinal study of early indirect and direct effects. *Journal of Experimental Child Psychology*, 164, 225–238. <https://doi.org/10.1016/j.jecp.2017.04.007>
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390–412. <https://doi.org/10.1016/j.jml.2007.12.00>
- Babayigit, S., Roulstone, S., & Wren, Y. (2020). Linguistic comprehension and narrative skills predict reading ability: A 9-year longitudinal study. *British Journal of Educational Psychology*, 91(1). <https://doi.org/10.1111/bjep.12353>
- Back, E., & Apperly, I. A. (2010). Two sources of evidence on the non-automaticity of true and false belief ascription. *Cognition*, 115(1), 54–70. <https://doi.org/10.1016/j.cognition.2009.11.008>

- Baerger, D. R., & McAdams, D. P. (1999). Life story coherence and its relation to psychological well-being. *Narrative inquiry*, 9(1), 69-96.
<https://doi.org/10.1075/ni.9.1.05bae>
- Baksh, R. A., Abrahams, S., Auyeung, B., & MacPherson, S. E. (2018). The Edinburgh Social Cognition Test (ESCoT): Examining the effects of age on a new measure of mentalising and social norm understanding. *PloS one*, 13(4), e0195818.
<https://doi.org/10.1371/jthenal.pone.0195818>
- Bambini, V., Van Looy, L., Demiddele, K., & Schaeken, W. (2021). What is the contribution of executive functions to communicative-pragmatic skills? Insights from aging and different types of pragmatic inference. *Cognitive Processing*, 22(3), 435–452. <https://doi.org/10.1007/s10339-021-01021-w>
- Baron-Cohen, S. (2000). *Theory of mind and autism: A review*. In G. Laraine Masters (Ed.), *International Review of Research in Mental Retardation* (Vol. Volume 23, pp. 169-184): Academic Press. [https://doi.org/10.1016/S0074-7750\(00\)80010-5](https://doi.org/10.1016/S0074-7750(00)80010-5)
- Baron-Cohen, S., Leslie, A. M., and Frith, U. (1985). Does the autistic child have a “mentalising”? *Cognition* 21, 37–46. [https://doi.org/10.1016/0010-0277\(85\)90022-8](https://doi.org/10.1016/0010-0277(85)90022-8)
- Baron-Cohen, S., Wheelwright, S., & Jolliffe, A. T. (1997). Is there a "language of the eyes"? Evidence from normal adults, and adults with autism or Asperger syndrome. *Visual cognition*, 4(3), 311-331. <https://doi.org/10.1080/713756761>
- Baron-Cohen, S., O'riordan, M., Stone, V., Jones, R., & Plaisted, K. (1999). Recognition of faux pas by normally developing children and children with Asperger syndrome or high-functioning autism. *Journal of autism and developmental disorders*, 29(5), 407-418. <https://doi.org/10.1023/A:1023035012436>
- Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The “Reading the Mind in the Eyes” Test revised version: a study with normal adults, and adults with

Asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 42(2), 241–251.

<https://doi.org/10.1017/S0021963001006643>

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

Baxter, L. (1992). Forms and functions of intimate play in personal relationships. *Human Communication*, 18, 336–363. <https://doi.org/10.1111/j.1468-2958.1992.tb00556.x>

Bazin, N., Brunet-Gouet, E., Bthedet, C., Kayser, N., Falissard, B., Hardy-Baylé, M. C., & Passerieux, C. (2009). Quantitative assessment of attribution of intentions to others in schizophrenia using an ecological video-based task: a comparison with manic and depressed patients. *Psychiatry research*, 167(1-2), 28-35.

<https://doi.org/10.1016/j.psychres.2007.12.010>

Bell, N. L., Lassiter, K. S., Matthews, T. D., & Hutchinson, M. B. (2001). Comparison of the peabody picture vocabulary test—Third edition and Wechsler adult intelligence scale—Third edition with university students. *Journal of clinical psychology*, 57(3), 417-422. <https://doi.org/10.1002/jclp.1024>

Benton A. L. (1968). Differential behavioural effects in frontal lobe disease.

Neuropsychologia 6, 53–60. DOI [https://doi.org/10.1016/0028-3932\(68\)90038-9](https://doi.org/10.1016/0028-3932(68)90038-9)

Bialecka-Pikul, M., Szpak, M., Zubek, J., Stępień-Nycz, M., Kołodziejczyk, A., Bosacki, S., & Homer, B. D. (2021). Measuring advanced mentalising: Do story-based tasks work? *Journal of Adolescence*, 93, 28–39.

<https://doi.org/10.1016/j.adolescence.2021.10.001>

- Bibby, H., & McDonald, S. (2005). Mentalising after traumatic brain injury. *Neuropsychologia*, 43(1), 99–114.
<https://doi.org/10.1016/j.neuropsychologia.2004.04.027>
- Binz, B., & Brune, M. (2010). Pragmatic language abilities, mentalising skills and executive functioning in schizophrenia spectrum disorders. *Clinical Neuropsychiatry: Journal of Treatment Evaluation*, 7(3), 91.
<https://link.gale.com/apps/doc/A315506614/HRC?u=anon~6191400&sid=googleScholar&xid=64a42b56>
- Bottiroli, S., Cavallini, E., Ceccato, I., Vecchi, T., & Lecce, S. (2016). Theory of Mind in aging: Comparing cognitive and affective components in the faux pas test. *Archives of Gerontology and Geriatrics*, 62, 152–162.
<https://doi.org/10.1016/j.archger.2015.09.009>
- Bischetti, L., Ceccato, I., Lecce, S., Cavallini, E., & Bambini, V. (2019). Pragmatics and mentalising in older adults' humor comprehension. *Current Psychology*, 42(19), 16191-16207. <https://doi.org/10.1007/s12144-019-00295-w>
- Bishop, D. V. (2014). *Uncommon Understanding (Classic Edition): Development and disorders of language comprehension in children*. Psychology Press.
- Bliksted, V., Videbech, P., Fagerlund, B., & Frith, C. (2017). The effect of positive symptoms on social cognition in first-episode schizophrenia is modified by the presence of negative symptoms. *Neuropsychology*, 31(2), 209–219.
<https://doi.org/10.1037/neu0000309>
- Bodner, K. E., Engelhardt, C. R., Minshew, N. J., & Williams, D. L. (2015). Making Inferences: Comprehension of Physical Causality, Intentionality, and Emotions in Discourse by High-Functioning Older Children, Adolescents, and Adults with

- Autism. *Journal of Autism and Developmental Disorders*, 45(9), 2721–2733.
<https://doi.org/10.1007/s10803-015-2436-3>
- Boerma, I. E., Mol, S. E., & Jolles, J. (2017). The role of home literacy environment, mentalizing, expressive verbal ability, and print exposure in third and fourth graders' reading comprehension. *Scientific Studies of Reading: The Official Journal of the Society for the Scientific Study of Reading*, 21(3), 179–193.
<https://doi.org/10.1080/10888438.2016.1277727>
- Bohn, M., & Frank, M. C. (2019). The pervasive role of pragmatics in early language. *Annual Review of Developmental Psychology*, 1, 223–249.
<https://doi.org/10.1146/annurev-devpsych-121318-085037>
- Booules-Katri, T.-M., Pedreño, C., Navarro, J.-B., Pamias, M., & Obiols, J. E. (2019). Mentalising (ToM) performance in high functioning autism (HFA) and schizotypal schizoid personality disorders (SSPD) patients. *Journal of Autism and Developmental Disorders*, 49(8), 3376–3386. <https://doi.org/10.1007/s10803-019-04058-1>
- Bora, E., Yucel, M., & Pantelis, C. (2009). Theory of mind impairment in schizophrenia: Meta-analysis. *Schizophrenia Research*, 109(1-3), 1–9.
<https://doi.org/10.1016/j.schres.2008.12.020>
- Borella, E., & de Ribaupierre, A. (2014). The role of working memory, inhibition, and processing speed in text comprehension in children. *Learning and Individual Differences*, 34, 86–92. <https://doi.org/10.1016/j.lindif.2014.05.001>
- Bosco, F. M., Gabbatore, I., Tirassa, M., & Testa, S. (2016). Psychometric Properties of the Mentalising Assessment Scale in a Sample of Adolescents and Adults. *Frontiers in Psychology*, 7, 566. <https://doi.org/10.3389/fpsyg.2016.00566>

- Bosco, F. M., Parola, A., Sacco, K., Zettin, M., & Angeleri, R. (2017). Communicative pragmatic disorders in traumatic brain injury: The role of mentalising and executive functions. *Brain and Language*, 168, 73–83.
<https://doi.org/10.1016/j.bandl.2017.01.007>
- Bosco, F. M., Gabbatore, I., Angeleri, R., Zettin, M., & Parola, A. (2018a). Do executive function and mentalising predict pragmatic abilities following traumatic brain injury? An analysis of sincere, deceitful and ironic communicative acts. *Journal of Communication Disorders*, 75, 102–117.
<https://doi.org/10.1016/j.jcomdis.2018.05.002>
- Bosco, F. M., Tirassa, M., & Gabbatore, I. (2018b). Why Pragmatics and Mentalising Do Not (Completely) Overlap. *Frontiers in Psychology*, 9, 1453.
<https://doi.org/10.3389/fpsyg.2018.01453>
- Brauer, M., & Curtin, J. J. (2018). Linear mixed-effects models and the analysis of non independent data: A unified framework to analyze categorical and continuous independent variables that vary within-subjects and/or within-items. *Psychological Methods*, 23(3), 389. <https://doi.org/10.1037/met0000159>
- Brewer, N., Young, R. L., & Barnett, E. (2017). Measuring mentalising in adults with autism spectrum disorder. *Journal of autism and developmental disorders*, 47(7), 1927-1941. <https://doi.org/10.1007/s10803-017-3080-x>
- Breheny, R., Ferguson, H. J., & Katsos, N. (2013). Investigating the time course of accessing conversational implicatures during incremental sentence interpretation. *Language and Cognitive Processes*, 28(4), 443-467.
<https://doi.org/10.1080/01690965.2011.649040>

- Brown, J. (2021). *Exploring the Relationship Between Socioeconomic Status, Language Exposure and Language Processing in Young Adults* (Doctoral dissertation, York St John University).
- Brüne, M. (2003). Mentalising and the role of IQ in chronic disorganized schizophrenia. *Schizophrenia research*, 60(1), 57-64. [https://doi.org/10.1016/S0920-9964\(02\)00162-7](https://doi.org/10.1016/S0920-9964(02)00162-7)
- Brüne, M., & Brüne-Cohrs, U. (2006). Mentalising—evolution, ontogeny, brain mechanisms and psychopathology. *Neuroscience & Biobehavioural Reviews*, 30(4), 437–455. <https://doi.org/10.1016/j.neubiorev.2005.08.001>
- Bruner, J. (1986). *Actual minds, possible words*. Harvard University Press.
- Brunet, E., Sarfati, Y., Hardy-Baylé, M. C., & Decety, J. (2000). A PET investigation of the attribution of intentions with a nonverbal task. *Neuroimage*, 11(2), 157-166. <https://doi.org/10.1006/nimg.1999.0525>
- Bryant, G. A., & Fox Tree, J. E. (2002). Recognizing verbal irony in spontaneous speech. *Metaphor and Symbol*, 17, 99-117. https://doi.org/10.1207/S15327868MS1702_2
- Burgers, C., Van Mulken, M., & Schellens, P. J. (2011). Finding irony: An introduction of the verbal irony procedure (VIP). *Metaphor and Symbol*, 26(3), 186-205. <https://doi.org/10.1080/10926488.2011.583194>
- Burgers, C., Van Mulken, M., & Schellens, P. J. (2012). Verbal irony: Differences in usage across written genres. *Journal of Language and Social Psychology*, 31(3), 290-310. <https://doi.org/10.1177/0261927X12446596>
- Burgers, C., Renardel de Lavalette, K. Y., & Steen, G. J. (2018). Metaphor, hyperbole, and irony: Uses in isolation and in combination in written discourse. *Journal of Pragmatics*, 127, 71–83. <https://doi.org/10.1016/j.pragma.2018.01.009>

- Byrne, R. W. (1996). Machiavellian intelligence. *Evolutionary Anthropology: Issues, News, and Reviews*, 5(5), 172–180. [https://doi.org/10.1002/\(sici\)1520-6505\(1996\)5:5%3C172::aid-evan6%3E3.0.co;2-h](https://doi.org/10.1002/(sici)1520-6505(1996)5:5%3C172::aid-evan6%3E3.0.co;2-h)
- Cabinio, M., Rossetto, F., Blasi, V., Savazzi, F., Castelli, I., Massaro, D., Valle, A., Nemni, R., Clerici, M., Marchetti, A., & Baglio, F. (2015). Mind-reading ability and structural connectivity changes in aging. *Frontiers in Psychology*, 6, 1808. <https://doi.org/https://doi.org/10.3389/fpsyg.2015.01808>
- Cain, K., & Oakhill, J. (2014). Reading comprehension and vocabulary: Is vocabulary more important for some aspects of comprehension?. *L'Annee psychologique*, 114(4), 647-662. <https://doi.org/10.4074/s0003503314004035>
- Cain, K., Oakhill, J. V., Barnes, M. A., & Bryant, P. E. (2001). Comprehension skill, inference-making ability, and their relation to knowledge. *Memory & Cognition*, 29(6), 850–859. <https://doi.org/10.3758/bf03196414>
- Cain, K., Oakhill, J. V., & Elbro, C. (2003). The ability to learn new word meanings from context by school-age children with and without language comprehension difficulties. *Journal of child language*, 30(3), 681-694. <https://doi.org/10.1017/s0305000903005713>
- Camodeca, A. (2019). Theory of Mind Performance in Broad Autism Phenotype Groups: Between-Group Differences and Predictor Variables. *Journal of Autism and Developmental Disorders*, 49(10), 4079–4096. <https://doi.org/10.1007/s10803-019-04126-6>
- Cane, J., Ferguson, H. J., & Apperly, I. A. (2017). Using perspective to resolve reference: The impact of cognitive load and motivation. *Journal of Experimental Psychology: Learning Memory and Cognition*, 43, 591–610. <https://doi.org/10.1037/xlm0000345>

- Carlson, S. M., & Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development*, 72(4), 1032–1053.
<https://doi.org/10.1111/1467-8624.00333>
- Cartwright, K. B. (2015). *Social understanding: the importance of mind reading for reading comprehension*. In *Executive skills and reading comprehension: a guide for educators* (pp. 201–228).
- Castelli, F., Happé, F., Frith, U., & Frith, C. (2000). Movement and Mind: A Functional Imaging Study of Perception and Interpretation of Complex Intentional Movement Patterns. *NeuroImage*, 12(3), 314–325. <https://doi.org/10.1006/nimg.2000.0612>
- Catts, H. W., Hogan, T. P., & Adlof, S. M. (2005). Developmental changes in reading and reading disabilities. In *The connections between language and reading disabilities* (pp. 38-51). Psychology Press. <https://doi.org/10.4324/9781410612052-9>
- Cavallini, E., Lecce, S., Bottiroli, S., Palladino, P., & Pagnin, A. (2013). Beyond false belief: mentalising in young, young-old, and old-old adults. *International Journal of Aging & Human Development*, 76(3), 181–198. <https://doi.org/10.2190/ag.76.3.a>
- Champagne-Lavau, M., & Stip, E. (2010). Pragmatic and executive dysfunction in schizophrenia. *Journal of Neurolinguistics*, 23(3), 285–296.
<https://doi.org/10.1016/j.jneuroling.2009.08.009>
- Cheang, H. S., & Pell, M. D. (2011). Recognizing sarcasm without language: A cross-linguistic study of English and Cantonese. *Pragmatics & Cognition*, 19(2), 203–223.
<https://doi.org/10.1075/pc.19.2.02che>
- Chen, K.-W., Lee, S.-C., Chiang, H.-Y., Syu, Y.-C., Yu, X.-X., & Hsieh, C.-L. (2017). Psychometric properties of three measures assessing advanced theory of mind: Evidence from people with schizophrenia. *Psychiatry Research*, 257, 490–496.
<https://doi.org/10.1016/j.psychres.2017.08.026>

- Chen, X., & Dewaele, J.-M. (2021). “We are not amused”. The appreciation of British humour by British and American English L1 users. *Language & Communication*, 79, 147–162. <https://doi.org/10.1016/j.langcom.2021.05.002>
- Chen, R. S., & Vellutino, F. R. (1997). Prediction of Reading Ability: A Cross-Validation Study of the Simple View of Reading. *Journal of Literacy Research*, 29(1), 1–24. <https://doi.org/10.1080/10862969709547947>
- Christopher, M. E., Miyake, A., Keenan, J. M., Pennington, B., DeFries, J. C., Wadsworth, S. J., Willcutt, E., & Olson, R. K. (2012). Predicting word reading and comprehension with executive function and speed measures across development: A latent variable analysis. *Journal of Experimental Psychology: General*, 141(3), 470–488. <https://doi.org/10.1037/a0027375>
- Clark, A. (1998). *Magic words: How language augments human cognition*. In P. Carruthers and J. Boucher (eds.), *Language and thought: Interdisciplinary themes* (pp. 162-183). Cambridge: Cambridge University Press. <https://doi.org/10.4324/9780203055069>
- Clark, H. H., & Lucy, P. (1975). Understanding what is meant from what is said: A study in conversationally conveyed requests. *Journal of verbal learning and verbal behavior*, 14(1), 56-72. [https://doi.org/10.1016/S0022-5371\(75\)80006-5](https://doi.org/10.1016/S0022-5371(75)80006-5)
- Clark, A., & Toribio, J. (2012). Magic words: how language augments human computation. In *Language and meaning in cognitive science* (pp. 33-51). Routledge. <https://era.ed.ac.uk/bitstream/handle/1842/1311/magic.pdf?sequence=1> <https://era.ed.ac.uk/bitstream/handle/1842/1311/magic.pdf?sequence=1>
- Cole, G. G., & Millett, A. C. (2019). The closing of the theory of mind: A critique of perspective-taking. *Psychonomic Bulletin & Review*, 26(6), 1787–1802. <https://doi.org/10.3758/s13423-019-01657-y>

- Compère, L., Mam-Lam-Fook, C., Amado, I., Nys, M., Lalanne, J., Grillon, M.-L., Bendjemaa, N., Krebs, M.-O., & Piolino, P. (2016). Self-reference recollection effect and its relation to mentalising: An investigation in healthy controls and schizophrenia. *Consciousness and Cognition*, 42, 51–64.
<https://doi.org/10.1016/j.concog.2016.03.004>
- Conners, F. A. (2008). Attentional control and the Simple View of reading. *Reading and Writing*, 22(5), 591–613. <https://doi.org/10.1007/s11145-008-9126-x>
- Corcoran, R., Mercer, G., and Frith, C. D. (1995). Schizophrenia symptomatology and social inference: investigating “mentalising” in people with schizophrenia. *Schizophr. Res.* 17, 5–13. [https://doi.org/10.1016/0920-9964\(95\)00024-G](https://doi.org/10.1016/0920-9964(95)00024-G)
- Corcoran, R., Cahill, C., & Frith, C. D. (1997). The appreciation of visual jokes in people with schizophrenia: a study of ‘mentalizing’ ability. *Schizophrenia research*, 24(3), 319-327. [https://doi.org/10.1016/S0920-9964\(96\)00117-X](https://doi.org/10.1016/S0920-9964(96)00117-X)Get rights and content
- Cunningham, A.E., & Stanovich, K.E. (1997). Early reading acquisition and its relation to reading experience and ability 10 years later. *Developmental Psychology*, 33(6), 934–945. <https://doi.org/10.1037/0012-1649.33.6.934>
- Currie, N. K., & Cain, K. (2015). Children’s inference generation: The role of vocabulary and working memory. *Journal of Experimental Child Psychology*, 137, 57–75.
<https://doi.org/10.1016/j.jecp.2015.03.005>
- Cutting, L. E., & Scarborough, H. S. (2006). Prediction of Reading Comprehension: Relative Contributions of Word Recognition, Language Proficiency, and Other Cognitive Skills Can Depend on How Comprehension Is Measured. *Scientific Studies of Reading*, 10(3), 277–299. https://doi.org/10.1207/s1532799xssr1003_5
- 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 Jong, P. F., & van der Leij, A. (2002). Effects of Phonological Abilities and Linguistic Comprehension on the Development of Reading. *Scientific Studies of Reading*, 6(1), 51–77. https://doi.org/10.1207/s1532799xssr0601_03
- De Mulder, H. N., Wijnen, F., & Coopmans, P. H. (2019). Interrelationships between Mentalising and language development: A longitudinal study of Dutch-speaking kindergartners. *Cognitive Development*, 51, 67-82.
<https://doi.org/10.1016/j.cogdev.2019.03.00>
- De Villiers, J. G. (1995). *Steps in the mastery of sentence complements*. Paper presented at the biennial meeting of the Society for Research in Child Development, Indianapolis, IN.
- De Villiers, J. (2000). Language and theory of mind: What are the developmental relationships?
- De Villiers, J. G., & De Villiers, P. A. (2000). Linguistic determinism and the understanding of false. *Children's reasoning and the mind*, 191.
https://scholarworks.umass.edu/aae_groundwork
- Deliens, G., Papastamou, F., Ruytenbeek, N., Geelhand, P., & Kissine, M. (2018). Selective Pragmatic Impairment in Autism Spectrum Disorder: Indirect Requests Versus Irony. *Journal of Autism and Developmental Disorders*, 48(9), 2938–2952.
<https://doi.org/10.1007/s10803-018-3561-6>
- Delis, D. C., Kaplan, E., & Kramer, J. H. (2001). Delis-Kaplan executive function system. *Assessment*. <https://doi.org/10.1080/09297040490911140>
- Demers, L. A., & Koven, N. S. (2015). The Relation of Alexithymic Traits to Affective Mentalising. *The American Journal of Psychology*, 128(1), 31–42.
<https://doi.org/10.5406/amerjpsyc.128.1.0031>

- Deska, J. C., & Hugenberg, K. (2017). The face-mind link: Why we see minds behind faces, and how others' minds change how we see their face. *Social and Personality Psychology Compass*, 11(12), e12361. <https://doi.org/10.1111/spc3.12361>
- Devellis, R. F. (2017). *Scale Development: Theory and Applications* (4th ed.). SAGE Publications.
- Devine, R. T., & Hughes, C. (2013). Silent films and strange stories: mentalising, gender, and social experiences in middle childhood. *Child Development*, 84(3), 989–1003. <https://doi.org/10.1111/cdev.12017>
- Devine, R. T., & Hughes, C. (2016). Measuring mentalising across middle childhood: Reliability and validity of the Silent Films and Strange Stories tasks. *Journal of Experimental Child Psychology*, 149, 23–40. <https://doi.org/10.1016/j.jecp.2015.07.011>
- Devine, R. T., & Hughes, C. (2019). Let's talk: Parents' mental talk (not mind-mindedness or mindreading capacity) predicts children's false belief understanding. *Child Development*, 90(4), 1236-1253. <https://doi.org/10.1111/cdev.12990>
- Dewart, H., & Summers, S. (1995). *The pragmatics profile of everyday communication skills in children*. Nfer-Nelson.
- Djikic, M., Oatley, K., Zoeterman, S., & Peterson, J. B. (2009). On being moved by art: How reading fiction transforms the self. *Creativity Research Journal*, 21(1), 24–29. <https://doi.org/10.1080/10400410802633392>
- Dockrell, J. E., Ricketts, J., Charman, T., & Lindsay, G. (2014). Exploring writing products in students with language impairments and autism spectrum disorders. *Learning and Instruction*, 32, 81-90. <https://doi.org/10.1016/j.learninstruc.2014.01.008>

- Dodell-Feder, D., Lincoln, S. H., Coulson, J. P., & Hooker, C. I. (2013). Using fiction to assess mental state understanding: a new task for assessing mentalising in adults. *PloS One*, 8(11), e81279. <https://doi.org/10.1371/journal.pone.0081279>
- Donaldson, M. (1987). *Children's minds*. London: Fontana Press, 1987.
<http://hdl.handle.net/11162/62747>
- Dore, R. A., Amend, S. J., Golinkoff, R. M., & Hirsh-Pasek, K. (2018). Mentalising: a Hidden Factor in Reading Comprehension? *Educational Psychology Review*, 30(3), 1067–1089. <https://doi.org/10.1007/s10648-018-9443-9>
- Dugdale, G., & Clark, C. (2008). *Literacy changes lives: An advocacy resource* (pp. 1-56). London: National Literacy Trust.
- Dumontheil, I., Apperly, I. A., & Blakemore, S.-J. (2010). Online usage of mentalising continues to develop in late adolescence. *Developmental Science*, 13(2), 331–338.
<https://doi.org/10.1111/j.1467-7687.2009.00888.x>
- Dziobek, I., Fleck, S., Kalbe, E., Rogers, K., Hassenstab, J., Brand, M., Kessler, J., Woike, J. K., Wolf, O. T., & Conit, A. (2006). Introducing MASC: a movie for the assessment of social cognition. *Journal of Autism and Developmental Disorders*, 36(5), 623–636. <https://doi.org/10.1007/s10803-006-0107-0>
- Ebert, S. (2020). Mentalising, language, and reading: Developmental relations from early childhood to early adolescence. *Journal of experimental child psychology*, 191, 104739. <https://doi.org/10.1016/j.jecp.2019.104739>
- Eddy, C. M., & Hansen, P. C. (2020). Predictors of performance on the Reading the Mind in the Eyes Test. *PLOS ONE*, 15(7), e0235529.
<https://doi.org/10.1371/journal.pone.0235529>
- Ekman, P., & Friesen, W. V. (1976). Measuring facial movement. *Environmental psychology and nonverbal behavior*, 1, 56-75. <https://doi.org/10.1007/bf01115465>

- Eloranta, A., Närhi, V. M., Eklund, K. M., Ahonen, T. P. S., & Aro, T. I. (2018). Resolving reading disability—Childhood predictors and adult-age outcomes. *Dyslexia*.
<https://doi.org/10.1002/dys.1605>
- Engelmann, J. B., & Pogosyan, M. (2013). Emotion perception across cultures: the role of cognitive mechanisms. *Frontiers in Psychology*, 4(118).
<https://doi.org/10.3389/fpsyg.2013.00118>
- Escalas, J. E. (2004). Imagine yourself in the product: Mental simulation, narrative transportation, and persuasion. *Journal of Advertising*, 33(2), 37–48.
<https://doi.org/10.1080/00913367.2004.10639163>
- Etchepare, A., Roux, S., Destailats, J. M., Cady, F., Fontanier, D., Couhet, G., & Prouteau, A. (2019). What are the specificities of social cognition in schizophrenia? A cluster-analytic study comparing schizophrenia with the general population. *Psychiatry Research*, 272, 369–379. <https://doi.org/10.1016/j.psychres.2018.12.042>
- Evans, D. K., & Hares, S. (2021). *Should Governments and Donors Prioritize Investments in Foundational Literacy and Numeracy?* (No. 579). Center for Global Development.
- Fairchild, S., & Papafragou, A. (2021). The Role of Executive Function and Theory of Mind in Pragmatic Computations. *Cognitive Science*, 45(2).
<https://doi.org/10.1111/cogs.12938>
- Fastame, M. C., & Carta, V. (2021). Do cognitive reserve and executive functions matter to perform the reading the mind in the eyes test in late lifespan? *Applied Neuropsychology: Adult*, 1-9. <https://doi.org/10.1080/23279095.2021.1972422>
- Federmeier K. D., Kutas M., Schul R. (2010). Age-related and individual differences in the use of prediction during language comprehension. *Brain Lang.* 149–161.
<https://doi.org/10.1016/j.bandl.2010.07.006>

- Federmeier, K. D., McLennan, D. B., De Ochoa, E., & Kutas, M. (2002). The impact of semantic memory organization and sentence context information on spoken language processing by younger and older adults: An ERP study. *Psychophysiology*, 39(2), 133-146. <https://doi.org/10.1111/1469-8986.3920133>
- Field, A. (2018). *Discovering Statistics Using IBM SPSS Statistics* (5th ed.). Sage Publications.
- Fernyhough, C. (2008). Getting Vygotskian about mentalising: Mediation, dialogue, and the development of social understanding. *Developmental review*, 28(2), 225-262. <https://doi.org/10.1016/j.dr.2007.03.001>
- Fernyhough, C., & Fradley, E. (2005). Private speech on an executive task: Relations with task difficulty and task performance. *Cognitive development*, 20(1), 103-120. <https://doi.org/10.1016/j.cogdev.2004.11.002>
- Fernyhough, C., & Meins, E. (2009). *Private speech and mentalising: Evidence for developing interfunctional relations*. <https://doi.org/10.1017/CBO9780511581533.008>
- Filiatrault-Veilleux, P., Bouchard, C., Trudeau, N., & Desmarais, C. (2016). Comprehension of inferences in a narrative in 3-to 6-year-old children. *Journal of Speech, Language, and Hearing Research*, 59(5), 1099-1110. https://doi.org/10.1044/2016_JSLHR-L-15-0252
- Filippova, E., & Astington, J. W. (2010). Children's understanding of social-cognitive and social-communicative aspects of discourse irony. *Child Development*, 81(3), 913-928. <https://doi.org/10.1111/j.1467-8624.2010.01442.x>
- Fisher, N., Happe, F., & Dunn, J. (2005). The relationship between vocabulary, grammar, and false belief task performance in children with autistic spectrum disorders and

- children with moderate learning difficulties. *Journal of Child Psychology and Psychiatry*, 46(4), 409–419. <https://doi.org/10.1111/j.1469-7610.2004.00371.x>
- Florit, E., & Cain, K. (2011). The Simple View of Reading: Is It Valid for Different Types of Alphabetic Orthographies? *Educational Psychology Review*, 23(4), 553–576. <https://doi.org/10.1007/s10648-011-9175-6>
- Ford, J.A., & Milosky, L. (2008). Inference generation during discourse and its relation to social competence: An online investigation of abilities of children with and without language impairment. *Journal of Speech, Language, and Hearing Research*, 51, 367–380. [https://doi.org/10.1044/1092-4388\(2008/027](https://doi.org/10.1044/1092-4388(2008/027)
- Franco, C. P., & Galvis, H. A. (2013). The role of situational context and linguistic context when testing EFL vocabulary knowledge in a language teacher education program: a preliminary approach. *Colombian Applied Linguistics Journal*, 15(1), 85-99. <https://doi.org/10.14483/udistrital.jthe.calj.2013.1.a06>
- Frank, C. K. (2018). Reviving pragmatic theory of mentalising. *AIMS Neuroscience*, 5(2), 116–131. <https://doi.org/10.3934/neuroscience.2018.2.116>
- Freed, J., Adams, C., & Lockton, E. (2015). Predictors of reading comprehension ability in primary school-aged children who have pragmatic language impairment. *Research in Developmental Disabilities*, 41-42, 13–21. <https://doi.org/10.1016/j.ridd.2015.03.003>
- Frith, C. D., & Allen, H. A. (1988). Language disorders in schizophrenia and their implications for neuropsychology. P. Bebbington, P. McGuffin (Eds.), *Schizophrenia: The major issues*, (pp. 172-186). Heinemann, Oxford.
- Frith, U., & Frith, C. D. (2003). Development and neurophysiology of mentalizing. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 358(1431), 459–473. <https://doi.org/10.1098/rstb.2002.1218>

- Frith, C. D., & Frith, U. (2006). The neural basis of mentalizing. *Neuron*, 50(4), 531–534.
<https://doi.org/10.1016/j.neuron.2006.05.001>
- Frith, U., & Frith, C. (2010). The social brain: allowing humans to boldly go where no other species has been. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1537), 165–176. <https://doi.org/10.1098/rstb.2009.0160>
- Frölander, H.-E., Möller, C., Rudner, M., Mishra, S., Marshall, J. D., Piacentini, H., & Lyxell, B. (2015). Theory-of-mind in individuals with Alström syndrome is related to executive functions, and verbal ability. *Frontiers in Psychology*, 6.
<https://doi.org/10.3389/fpsyg.2015.01426>
- Fung, H. H., Isaacowitz, D. M., Lu, A. Y., Wadlinger, H. A., Goren, D., & Wilson, H. R. (2008). Age-related positivity enhancement is not universal: Older Chinese look away from positive stimuli. *Psychology and Aging*, 23(2), 440–446.
<https://doi.org/10.1037/0882-7974.23.2.440>
- Gallagher, H. L., Happé, F., Brunswick, N., Fletcher, P. C., Frith, U., & Frith, C. D. (2000). Reading the mind in cartoons and stories: an fMRI study of ‘mentalising’ in verbal and nonverbal tasks. *Neuropsychologia*, 38(1), 11-21. [https://doi.org/10.1016/S0028-3932\(99\)00053-6](https://doi.org/10.1016/S0028-3932(99)00053-6)
- García, J. R., & Cain, K. (2014). Decoding and reading comprehension: A meta-analysis to identify which reader and assessment characteristics influence the strength of the relationship in English. *Review of Educational Research*, 84(1), 74-111.
<https://doi.org/10.3102/0034654313499616>
- Gavilán, J. M., & García-Albea, J. E. (2011). Mentalising and language comprehension in schizophrenia: Poor mindreading affects figurative language comprehension beyond intelligence deficits. *Journal of Neurolinguistics*, 24(1), 54–69.
<https://doi.org/10.1016/j.jneuroling.2010.07.006>

- Georgiou, G. K., Das, J. P., & Hayward, D. (2008). Revisiting the “Simple View of Reading” in a Group of Children With Poor Reading Comprehension. *Journal of Learning Disabilities*, 42(1), 76–84. <https://doi.org/10.1177/0022219408326210>
- Geurts, B. (2010). *Quantity implicatures*. Cambridge University Press.
<https://doi.org/10.1017/CBO9780511975158>
- Gelman, A., & Hill, J., (2006). *Data Analysis Using Regression and Multilevel/Hierarchical Models*, Cambridge University Press.
- Goddard, C. (2018). “Joking, kidding, teasing”: Slippery categories for cross-cultural comparison but key words for understanding Anglo conversational humour. *Intercultural Pragmatics*, 15(4), 487-514. <https://doi.org/10.1515/ip-2018-0017>
- Gibbons, Z. C., Snowden, J. S., Thompson, J. C., Happé, F., Richardson, A., & Neary, D. (2007). Inferring thought and action in motor neurone disease. *Neuropsychologia*, 45(6), 1196–1207. <https://doi.org/10.1016/j.neuropsychologia.2006.10.008>
- Gibbs Jr, R. W. (1981). Your wish is my command: Convention and context in interpreting indirect requests. *Journal of Verbal Learning and Verbal Behaviour*, 20(4), 431-444.
[https://doi.org/10.1016/S0022-5371\(81\)90542-9](https://doi.org/10.1016/S0022-5371(81)90542-9)
- Gibbs, R. W., Jr. (1987). Mutual knowledge and the psychology of conversational inference. *Journal of Pragmatics*, 11(5), 561–588 [https://doi.org/10.1016/0378-2166\(87\)90180-9](https://doi.org/10.1016/0378-2166(87)90180-9).
- Gibbs R. W. 1994. *The poetics of mind: Figurative thought, language, and understanding*. New York: Cambridge University Press. <https://doi.org/10.1075/pc.3.1.15tur>
- Gibbs, R. W., Jr. (2000). Irony in talk among friends. *Metaphor and Symbol*, 15, 5-27.
<https://doi.org/10.1080/10926488.2000.9678862>
- Gibbs, R. W. (1984). Literal Meaning and Psychological Theory. *Cognitive Science*, 8(3), 275–304. https://doi.org/10.1207/s15516709cog0803_4

- Gibbs, R., & Colston, H. (2007). *The future of irony studies*. In R. Gibbs & H. Colston (Eds.), *Irony in language and thought: A cognitive science reader* (pp. 581–595). New York: Erlbaum.
- Gibbs, R. W., & Colston, H. L. (2020). Pragmatics Always Matters: An Expanded Vision of Experimental Pragmatics. *Frontiers in Psychology, 11*.
<https://doi.org/10.3389/fpsyg.2020.01619>
- Gildea, P., & Glucksberg, S. (1983). On understanding metaphor: The role of context. *Journal of verbal learning and verbal behavior, 22*(5), 577-590.
[https://doi.org/10.1016/S0022-5371\(83\)90355-9](https://doi.org/10.1016/S0022-5371(83)90355-9)
- Gillam, R. B., & Pearson, N. A. (2004). *Test of Narrative Language*. Austin, TX: Pro-Ed.
- Giordano, M., Jalil Rasgado-Toledo, Averi Giudicessi, & Lizcano-Cortés, F. (2020). Theory of Mind and its elusive structural substrate. *OSF Preprints* (OSF Preprints).
<https://doi.org/10.31219/osf.io/5pf2h>
- Giovagnoli, A. R. (2019). Mentalising across lifespan from ages 16 to 81 years. *Epilepsy & Behavior: E&B, 100*, 106349. <https://doi.org/10.1016/j.yebeh.2019.05.044>
- Goel, V., Grafman, J., Sadato, N., & Hallett, M. (1995). Modelling other minds. *NeuroReport-International Journal for Rapid Communications of Research in Neuroscience, 6*(13), 1741-1746. <https://doi.org/10.1097/00001756-199509000-00009>
- Golan, O., Baron-Cohen, S., Hill, J.J., & Golan, Y. (2006). The ‘reading the mind in films’ task: Complex emotion recognition in adults with and without autism spectrum conditions. *Social Neuroscience, 1*, 111–123.
<https://doi.org/10.1080/17470910600980986>
- Golan, O., Baron-Cohen, S., Hill, J. J., & Golan, Y. (2006). The “reading the mind in films” task: complex emotion recognition in adults with and without autism spectrum

conditions. *Social Neuroscience*, 1(2), 111-123.

<https://doi.org/10.1080/17470910600980986>

Grainger, S. A., Henry, J. D., Naughtin, C. K., Comino, M. S., & Dux, P. E. (2018). Implicit false belief tracking is preserved in late adulthood. *Quarterly Journal of Experimental Psychology*, 71(9), 1980–1987.

<https://doi.org/10.1177/1747021817734690>

Grazzani, I., & Ornaghi, V. (2012). How do use and comprehension of mental-state language relate to theory of mind in middle childhood? *Cognitive Development*, 27(2), 99–111. <https://doi.org/10.1016/j.cogdev.2012.03.002>

Grice, H. P. (1989). *Studies in the Way of Words*. Cambridge, MA: Harvard University Press.

Grimm, J., Grimm, W., ZIPES, J., & DEZSÖ, A. (2014). *The Original Folk and Fairy Tales of the Brothers Grimm: The Complete First Edition*. Princeton University Press.

<http://www.jstor.org/stable/j.ctt6wq18v>

Grodner, D., & Sedivy, J. C. (2011). 10 The Effect of Speaker-Specific Information on Pragmatic Inferences. In *The processing and acquisition of reference* (Vol. 2327, pp. 239-272). MIT Press.

Haberlandt, K. F., & Graesser, A. C. (1985). Component processes in text comprehension and some of their interactions. *Journal of Experimental Psychology: General*, 114, 357–374. <http://dx.doi.org/10.1037/0096-3445.114.3.357>

Hamilton, L. G., O'Halloran, I., & Cutting, N. (2020). Individual differences in narrative production in late childhood: Associations with age and fiction reading experience. *First Language*, 0142723720946553. <https://doi.org/10.1177/0142723720946553>

- Hao, J., & Liu, Y. (2016). A Mind-Reader Does Not Always Have Deontological Moral Judgments and Prosocial Behaviour: A Developmental Perspective. *Frontiers in Psychology*, 7, 1261. <https://doi.org/10.3389/fpsyg.2016.01261>
- Happé, F. G. (1993). Communicative competence and mentalising in autism: A test of relevance theory. *Cognition*, 48(2), 101–119. [https://doi.org/10.1016/0010-0277\(93\)90026-r](https://doi.org/10.1016/0010-0277(93)90026-r)
- Happé, F. G. (1994). An advanced test of mentalising: Understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *Journal of autism and Developmental disorders*, 24(2), 129-154. <https://doi.org/10.1007/BF02172093>
- Happé, F. G., Winner, E., & Brownell, H. (1998). The getting of wisdom: mentalising in old age. *Developmental psychology*, 34(2), 358. <https://doi.org/10.1037/0012-1649.34.2.358>
- Happé, F., Brownell, H., & Winner, E. (1999). Acquired mentalising impairments following stroke. *Cognition*, 70(3), 211-240. [https://doi.org/10.1016/S0010-0277\(99\)00005-0](https://doi.org/10.1016/S0010-0277(99)00005-0)
- Harkness, K. L., Jacobson, J. A., Duong, D., & Sabbagh, M. A. (2010). Mental state decoding in past major depression: Effect of sad versus happy mood induction. *Cognition & Emotion*, 24(3), 497–513. <https://doi.org/10.1080/02699930902750249>
- Harris, L.T., (2017). *Invisible mind: flexible social cognition and dehumanization*. Cambridge, MA: MIT Press. <http://10.7551/mitpress/9780262035965.001.0001>
- Hartman, D. E. (2009). Wechsler Adult Intelligence Scale IV (WAIS IV): return of the gold standard. *Applied neuropsychology*, 16(1), 85-87. <https://doi.org/10.1080/09084280802644466>

- Harvey, A., Spicer-Cain, H., Botting, N., & Henry, L. (2024). Narrative abilities of autistic and non-autistic adolescents: The role of mentalising and executive function. *Autism Research*. <https://doi.org/10.1002/aur.3272>
- Hayward, E. O., and Homer, B. D. (2017). Reliability and validity of advanced mentalising measures in middle childhood and adolescence. *Br. J. Dev. Psychol.* 35, 454–462. <https://doi.org/10.1111/bjdp.12186>
- Heider, F., & Simmel, M. (1944). An experimental study of apparent behaviour. *The American Journal of Psychology*, 57, 243–259. <https://doi.org/10.2307/1416950>
- Hernandez, D. J., & Casey, A. E. (2120). *Double Jeopardy: How Third-Grade Reading Skills and Poverty Influence High School Graduation*. Annie E. Casey Foundation. 701 Saint Paul Street, Baltimore, Md 2. Tel: 410-547-; Fax: 410-547-; Web Site: <Http://Www.Aecf.org>.
- Henry, J. D., Phillips, L. H., Ruffman, T., & Bailey, P. E. (2013). A meta-analytic review of age differences in mentalising. *Psychology and Aging*, 28(3), 826. <https://doi.org/10.1037/a0030677>
- Herbet, G., Lafargue, G., Bonnetblanc, F., Moritz-Gasser, S., Menjot de Champfleur, N., & Duffau, H. (2014). Inferring a dual-stream model of mentalizing from associative white matter fibres disconnection. *Brain: A Journal of Neurology*, 137(Pt 3), 944–959. <https://doi.org/10.1093/brain/awt370>
- Heyes, C. M., & Frith, C. D. (2014). The cultural evolution of mind reading. *Science*, 344(6190), 1243091. <https://doi.org/10.1126/science.1243091>
- Hezel, D. M., & McNally, R. J. (2014). Theory of Mind Impairments in Social Anxiety Disorder. *Behavior Therapy*, 45(4), 530–540. <https://doi.org/10.1016/j.beth.2014.02.010>

- Hirotsani, M., Frazier, L., & Rayner, K. (2006). Punctuation and intonation effects on clause and sentence wrap-up: Evidence from eye movements. *Journal of Memory and Language*, 54(3), 425-443. <https://doi.org/10.1016/j.jml.2005.12.001>
- Hirshorn E. A., Thompson-Schill S. L. (2006). Role of the left inferior frontal gyrus in covert word retrieval: neural correlates of switching during verbal fluency. *Neuropsychologia* 44, 2547–2557. <https://doi.org/10.1016/j.neuropsychologia.2006.03.035>
- Hjetland, H. N., Lervåg, A., Lyster, S.-A. H., Hagtvet, B. E., Hulme, C., & Melby-Lervåg, M. (2019). Pathways to reading comprehension: A longitudinal study from 4 to 9 years of age. *Journal of Educational Psychology*, 111(5), 751–763. <https://doi.org/10.1037/edu0000321>
- Ho, C. S.-H., Chow, B. W.-Y., Wong, S. W.-L., Waye, M. M. Y., & Bishop, D. V. M. (2012). The Genetic and Environmental Foundation of the Simple View of Reading in Chinese. *PLoS ONE*, 7(10), e47872. <https://doi.org/10.1371/journal.pone.0047872>
- Ho, K. K. Y., Lui, S. S. Y., Hung, K. S. Y., Wang, Y., Li, Z., Cheung, E. F. C., & Chan, R. C. K. (2015). Theory of mind impairments in patients with first-episode schizophrenia and their unaffected siblings. *Schizophrenia Research*, 166(1-3), 1–8. <https://doi.org/10.1016/j.schres.2015.05.033>
- Hochstein, L., Bale, A., & Barner, D. (2018). Scalar implicature in absence of epistemic reasoning? The case of autism spectrum disorder. *Language Learning and Development*, 14(3), 224-240. <https://doi.org/10.1080/15475441.2017.1343670>
- Hodapp, A. F., & Gerken, K. C. (1999). Correlations between scores for Peabody picture vocabulary test—III and the Wechsler intelligence scale for children—III. *Psychological Reports*, 84(3_suppl), 1139-1142. <https://doi.org/10.2466/pr0.1999.84.3c.1139>

- Hoicka, E. (2014). The pragmatic development of humor. *Pragmatic development in first language acquisition*, 10, 219. <https://doi.org/10.1075/tilar.10.13hoi>
- Hoicka, E., & Gattis, M. (2008). Do the wrong thing: How toddlers tell a joke from a mistake. *Cognitive Development*, 23(1), 180-190.
<https://doi.org/10.1016/j.cogdev.2007.06.001>
- Hoicka, E., Jutsum, S., & Gattis, M. (2008). Humor, abstraction, and disbelief. *Cognitive Science*, 32(6), 985-1002. <https://doi.org/10.1080/03640210801981841>
- Hogrefe, G.-J., Wimmer, H., & Perner, J. (1986). Ignorance versus False Belief: A Developmental Lag in Attribution of Epistemic States. *Child Development*, 57(3), 567. <https://doi.org/10.2307/1130337>
- Holopainen, E. (2003). *Understanding strategies and difficulties in reading and reading text in the third and ninth grades of primary school* (No. 218). University of Jyväskylä.
- Honan, C. A., McDonald, S., Gowland, A., Fisher, A., & Randall, R. K. (2015). Deficits in comprehension of speech acts after TBI: The role of mentalising and executive function. *Brain and Language*, 150, 69–79.
<https://doi.org/10.1016/j.bandl.2015.08.007>
- Hoover, W. A. (2023). The simple view of reading and its broad types of reading difficulties. *Reading and Writing*, 1-22. <https://doi.org/10.1007/s11145-023-10471-x>
- Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and writing*, 2(2), 127-160. <https://doi.org/10.1007/bf00401799>
- Horan, W. P., Nuechterlein, K. H., Wynn, J. K., Lee, J., Castelli, F., & Green, M. F. (2008). Disturbances in the spontaneous attribution of social meaning in schizophrenia. *Psychological Medicine*, 39(4), 635–643.
<https://doi.org/10.1017/s0033291708003838>

- Horton, W. S., & Brennan, S. E. (2016). The Role of Metarepresentation in the Production and Resolution of Referring Expressions. *Frontiers in Psychology*, 7.
<https://doi.org/10.3389/fpsyg.2016.01111>
- Horton, W. S. & Grieg, R. J. (2005). Conversational common ground and memory processes in language production. *Discourse Processes*, 40(1), 1-35.
https://doi.org/10.1207/s15326950dp4001_1
- Hughes, C. (2011). *Social understanding and social lives: From toddlerhood through to the transition to school*. Psychology Press. <https://doi.org/10.1002/icd.760>
- Hughes, C., & Devine, R. T. (2015). Individual differences in mentalising from preschool to adolescence: Achievements and directions. *Child Development Perspectives*, 9(3), 149–153. <https://doi.org/10.1111/cdep.12124>
- Hughes, C., Jaffee, S. R., Happé, F., Taylor, A., Caspi, A., & Moffitt, T. E. (2005). Origins of individual differences in mentalising: from nature to nurture? *Child Development*, 76(2), 356–370. https://doi.org/10.1111/j.1467-8624.2005.00850_a.x
- Hulme, C., & Snowling, M. J. (2014). The interface between spoken and written language: developmental disorders. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369 (1634), <https://doi.org/10.1098/rstb.2012.0395>
- IBM Corp. (2020). IBM SPSS Statistics for Windows (Version 27.0) [Computer software]. IBM Corp.
- Jackson, S. (2022). Exploring the mutually reinforcing relationship between mentalising and reading in adult readers. *Journal of Research in Reading*, 45(2), 189-203.
<https://doi.org/10.1111/1467->

- Jenkins, J. M., & Astington, J. W. (1996). Cognitive factors and family structure associated with theory of mind development in young children. *Developmental Psychology*, 31(1), 70–78. <https://doi.org/10.1037/0012-1649.32.1.70>.
- Johnson, A. H., & Barrett, J. (2003). The Role of Control in Attributing Intentional Agency to Inanimate Objects. *Journal of Cognition and Culture*, 3(3), 208–217. <https://doi.org/10.1111/1467-9817.12393>
- Johnston, T. C., & Kirby, J. R. (2006). The Contribution of Naming Speed to the Simple View of Reading. *Reading and Writing*, 19(4), 339–361. <https://doi.org/10.1007/s11145-005-4644-2>
- Johnson-Laird, P. N. (1983). *Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness*. Cambridge, MA: Harvard University Press.
- Joshi, R. M., & Aaron, P. G. (2000). The component model of reading: Simple view of reading made a little more complex. *Reading Psychology*, 21(2), 85-97. [*https://doi.org/10.1080/02702710050084428](https://doi.org/10.1080/02702710050084428)
- Joyal, M., Bonneau, A., & Fecteau, S. (2016). Speech and language therapies to improve pragmatics and discourse skills in patients with schizophrenia. *Psychiatry research*, 240, 88-95. <https://doi.org/10.1016/j.psychres.2016.04.010>
- Juel, C. (1988). Learning to read and write: A longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology*, 80(4), 437–447. <https://doi.org/10.1037/0022-0663.80.4.437>
- Katz, A. N., & Ferretti, T. R., (2001). Moment-by-moment reading of proverbs in literal and nonliteral contexts. *Metaphor and Symbol* 16. 193- 221. <https://doi.org/10.1080/10926488.2001.9678895>

- Katz, L., Brancazio, L., Irwin, J., Katz, S., Magnuson, J., & Whalen, D. H. (2012). What lexical decision and naming tell us about reading. *Reading and Writing*, 25(6), 1259–1282. <https://doi.org/10.1007/s11145-011-9316-9>
- Keltner, D., Young, R., Heerey, E., Oeming, C., & Monarch, N. (1998). Teasing in hierarchical and intimate relations. *Journal of Personality and Social Psychology*, 75, 1231–1247. <https://doi.org/10.1037/0022-3514.75.5.1231>
- Kendeou, P., van den Broek, P., White, M. J., & Lynch, J. S. (2009). Predicting reading comprehension in early elementary school: The independent contributions of oral language and decoding skills. *Journal of Educational Psychology*, 101(4), 765–778. <https://doi.org/10.1037/a0015956>
- Kendeou, P., van den Broek, P., Helder, A., & Karlsson, J. (2014). A Cognitive View of Reading Comprehension: Implications for Reading Difficulties. *Learning Disabilities Research & Practice*, 29(1), 10–16. <https://doi.org/10.1111/ldrp.12025>
- Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking Perspective in Conversation: The Role of Mutual Knowledge in Comprehension. *Psychological Science*, 11(1), 32–38. <https://doi.org/10.1111/1467-9280.00211>
- Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on mentalising use in adults. *Cognition*, 89, 25–41. [https://doi.org/10.1016/S0010-0277\(03\)00064-7](https://doi.org/10.1016/S0010-0277(03)00064-7)
- Khorashad, B. S., Baron-Cohen, S., Roshan, G. M., Kazemian, M., Khazai, L., Aghili, Z., et al. (2015). The “Reading the Mind in the Eyes” test: investigation of psychometric properties and test–retest reliability of the Persian version. *Journal of Autism and Developmental Disorders* 45, 2651–2666. <https://doi.org/10.1007/s10803-015-2427-4>
- Kidd, D. C., & Castano, E. (2013). Reading literary fiction improves mentalising. *Science*, 342(6156), 377–380. <https://doi.org/10.1126/science.1239918>

- Kidd, D., & Castano, E. (2019). Reading Literary Fiction and Mentalising: Three Preregistered Replications and Extensions of Kidd and Castano (2013). *Social Psychological and Personality Science*, 10(4), 522–531.
<https://doi.org/10.1177/1948550618775410>
- Kidd, D. C., & Castano, E. (2017). Panero et al. (2016): Failure to replicate methods caused the failure to replicate results. *Journal of Personality and Social Psychology*, 112(3), e1–e4. <https://doi.org/10.1037/pspa0000072>
- Kidd, D., & Castano, E. (2019). Reading literary fiction and theory of mind: Three preregistered replications and extensions of Kidd and Castano (2013). *Social Psychological and Personality Science*, 10(4), 522–531.
<https://doi.org/10.1177/19485506187754>
- Kim, Y.-S. (2015). Language and cognitive predictors of text comprehension: Evidence from multivariate analysis. *Child Development*, 86, 128–144.
<https://doi.org/10.1111/cdev.12293>
- Kim, Y.-S. G. (2016). Direct and mediated effects of language and cognitive skills on comprehension or oral narrative texts (listening comprehension) for children. *Journal of Experimental Child Psychology*, 141, 101–120. <https://doi.org/10.1016/j.jecp.2015.08.003>
- Kim, Y.-S. G. (2017). Why the Simple View of Reading Is Not Simplistic: Unpacking Component Skills of Reading Using a Direct and Indirect Effect Model of Reading (DIER). *Scientific Studies of Reading: The Official Journal of the Society for the Scientific Study of Reading*, 21(4), 310–333.
<https://doi.org/10.1080/10888438.2017.1291643>

- Kim, Y. S. G., & Petscher, Y. (2020). Influences of individual, text, and assessment factors on text/discourse comprehension in oral language (listening comprehension). *Annals of Dyslexia*, 1-20. <https://doi.org/10.1007/s11881-020-00208-8>
- Kim, Y.-S. G., Dore, R., Cho, M., Golinkoff, R., & Amend, S. J. (2021). Mentalising, mental state talk, and discourse comprehension: Mentalising process is more important for narrative comprehension than for informational text comprehension. *Journal of Experimental Child Psychology*, 209, 105181. <https://doi.org/10.1016/j.jecp.2021.105181>
- Kintsch, W. (1988). The role of knowledge in discourse comprehension: a construction integration model. *Psychological Review*, 95(2), 163–182. <https://doi.org/10.1037/0033-295x.95.2.163>
- Kirby, J. R., & Savage, R. S. (2008). Can the simple view deal with the complexities of reading? *Literacy*, 42(2), 75–82. <https://doi.org/10.1111/j.1741-4369.2008.00487.x>
- Kirkland, R. A., Baker, C. A., Johnson, C., Peterson, E., & Potos, S. (2012). Meta-analysis reveals a moderate relationship between tests of mentalising and the Eyes Test. *Poster Presented at the Association of Psychological Science (Chicago):*
- Konrath, S., Corneille, O., Bushman, B. J., and Luminet, O. (2014). The relationship between narcissistic exploitativeness, dispositional empathy, and emotion recognition abilities. *J. Nonverbal Behaviour*. 38, 129–143. <https://doi.org/10.1007/s10919-013-0164-y>
- Kovacs, A. M., Teglas, E., & Endress, A. D. (2010). The Social Sense: Susceptibility to Others' Beliefs in Human Infants and Adults. *Science*, 330(6012), 1830–1834. <https://doi.org/10.1126/science.1190792>
- Krenzke, T., Mohadjer, L., Li, J., Van de Kerckhove, W., Li, L., Ren, W., & Adbaru, H. (2018). Programme for the international assessment of adult competencies.

<https://www.oecd.org/skills/piaac/publications/workingpapers/PIAACSAEInitialResearchReport10Sept2018.pdf>

Kristen, S., Rossmann, F., & Sodian, B. (2014). Theory of own mind and autobiographical memory in adults with ASD. *Research in Autism Spectrum Disorders*, 8(7), 827–837. <https://doi.org/10.1016/j.rasd.2014.03.009>

Marocchini, E., & Domaneschi, F. (2022). “Can you read my mind?” Conventionalised indirect requests and Theory of Mind abilities. *Journal of Pragmatics*, 193, 201–221. <https://doi.org/10.1016/j.pragma.2022.03.011>

Ladegaard, N., Larsen, E. R., Videbech, P., & Lysaker, P. H. (2014). Higher-order social cognition in first-episode major depression. *Psychiatry Research*, 216(1), 37–43. <https://doi.org/10.1016/j.psychres.2013.12.010>

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2005). *International affective picture system (IAPS): Affective ratings of pictures and instruction manual* (pp. A-8). Gainesville, FL: NIMH, Center for the Study of Emotion & Attention. <https://doi.org/10.3758/brm.40.2.512>

Langdon, R., & Coltheart, M. (1999). Mentalising, schizotypy, and schizophrenia. *Cognition*, 71(1), 43–71. [https://doi.org/10.1016/s0010-0277\(99\)00018-9](https://doi.org/10.1016/s0010-0277(99)00018-9)

Langdon, R., Connors, M. H., Still, M., Ward, P. B., & Catts, S. (2014). Mentalising and neurocognition in early psychosis: a quasi-experimental study. *BMC psychiatry*, 14(1), 1–7. <https://doi.org/10.1186/s12888-014-0316-6>

Language, & Consortium, R. R. (2015). Learning to read: Should we keep things simple? *Reading Research Quarterly*, 50(2), 151–169. <https://doi.org/10.1002/rrq.99>

Language and Reading Research Conso, & Muijselaar, M. M. L. (2018). The dimensionality of inference making: Are local and global inferences distinguishable? *Scientific*

- Studies of Reading: The Official Journal of the Society for the Scientific Study of Reading*, 22(2), 117–136. <https://doi.org/10.1080/10888438.2017.1371179>
- Laumann Long, Raymond J. Shaw, L. (2000). Adult age differences in vocabulary acquisition. *Educational Gerontology*, 26(7), 651–664.
<https://doi.org/10.1080/03601270050200644>
- Lecce, S., Ceccato, I., & Cavallini, E. (2019). Mentalising, mental state talk and social relationships in aging: the case of friendship. *Aging & mental health*, 23(9), 1105–1112. <https://doi.org/10.1080/13607863.2018.1479832>
- Lecce, S., & Bianco, F. (2018). The role of false-belief understanding in preschoolers' development of metamemory: A training study. *European Journal of Developmental Psychology*, 16(6), 697–711. <https://doi.org/10.1080/17405629.2018.1496908>
- Lecce, S., Federica B., Devine, R. T., and Hughes, C. (2017). "Relations between mentalising and executive function in middle childhood: A short-term longitudinal study." *Journal of Experimental Child Psychology*, 163, 69–86.
<https://doi.org/10.1016/j.jecp.2017.06.011>
- Lecce, S., Ronchi, L., Del Sette, P., Bischetti, L., & Bambini, V. (2019). Interpreting physical and mental metaphors: Is Mentalising associated with pragmatics in middle childhood? *Journal of child language*, 46(2), 393–407.
<https://doi.org/10.1017/s030500091800048x>
- Lecciso, F., Levante, A., Baruffaldi, F., & Petrocchi, S. (2016). Mentalising in deaf adults. *Cogent Psychology*, 3(1), 1264127.
<https://doi.org/10.1080/23311908.2016.1264127>
- Lee TMC, Ip AKY, Wang K, Xi C, Hu P, Mak HKF, Han SH, Chan CCH (2010). Faux pas deficits in people with medial frontal lesions as related to impaired understanding of a speaker's mental state. *Neuropsychologia.*, 48 (6):1670–

1676. <https://doi.org/10.1016/j.neuropsychologia.2010.02.012>.
- Lee, H.-S., Corbera, S., Poltorak, A., Park, K., Assaf, M., Bell, M. D., Wexler, B. E., Cho, Y.-I., Jung, S., Brocke, S., & Choi, K.-H. (2018). Measuring mentalising in schizophrenia research: Cross-cultural validation. *Schizophrenia Research*, 201, 187–195. <https://doi.org/10.1016/j.schres.2018.06.022>
- Leinonen, E., & Letts, C. (1997). Referential communication tasks: performance by normal and pragmatically impaired children. *International Journal of Language & Communication Disorders*, 32(2s), 53-65. <https://doi.org/10.1111/j.1460-6984.1997.tb01624.x>
- Lemhöfer, K., Huestegge, L., & Mulder, K. (2018). Another cup of TEE? The processing of second language near-cognates in first language reading. *Language, Cognition and Neuroscience*, 33(8), 968-991. <https://doi.org/10.1080/23273798.2018.1433863>
- Leslie, L., & Caldwell, J. S. (2011). Qualitative reading inventory (5th ed.). Boston, MA: Pearson.
- Leslie, A. M., & Thaiss, L. (1992). Domain specificity in conceptual development: Neuropsychological evidence from autism. *Cognition*, 43(3), 225–251. [https://doi.org/10.1016/0010-0277\(92\)90013-8](https://doi.org/10.1016/0010-0277(92)90013-8)
- Li, X., Wang, K., Wang, F., Tao, Q., Xie, Y., & Cheng, Q. (2013). Aging of mentalising: the influence of educational level and cognitive processing. *International Journal of Psychology: Journal International de Psychologie*, 48(4), 715–727. <https://doi.org/10.1080/00207594.2012.673724>
- Lichtenstein, E. H., & Brewer, W. F. (1980). Memory for goal-directed events. *Cognitive Psychology*, 12(3), 412-445. [https://doi.org/10.1016/0010-0285\(80\)90015-8](https://doi.org/10.1016/0010-0285(80)90015-8)
- Lonigan, C. J., Burgess, S. R., & Anthony, J. L. (2000). Development of emergent literacy and early reading skills in preschool children: Evidence from a latent-variable

longitudinal study. *Developmental Psychology*, 36(5), 596–613.

<https://doi.org/10.1037/0012-1649.36.5.596>

Lonigan, C. J., Burgess, S. R., & Schatschneider, C. (2018). Examining the Simple View of Reading With Elementary School Children: Still Simple After All These Years. *Remedial and Special Education*, 39(5), 260–273.

<https://doi.org/10.1177/0741932518764833>

Long, E. L., Cuve, H. C., Conway, J., Catmur, C., & Bird, G. (2022). The Interview Task: Novel Mentalising Task Demonstrates Representation of Minds in Mental State Inference. <https://doi.org/10.1038/s41598-022-25490-x>

Long, M. R., Horton, W. S., Rohde, H., & Sorace, A. (2018). Individual differences in switching and inhibition predict perspective-taking across the lifespan. *Cognition*, 170, 25-30. <https://doi.org/10.1016/j.cognition.2017.09.004>

Lough, S., Kipps, C. M., Treise, C., Watson, P., Blair, J. R., & Hodges, J. R. (2006). Social reasoning, emotion and empathy in frontotemporal dementia. *Neuropsychologia*, 44(6), 950–958. <https://doi.org/10.1016/j.neuropsychologia.2005.08.009>

Lin, DeWayne Moore, Karen M. Zabruc, L.-M. (2000). Metacomprehension knowledge and comprehension of expository and narrative texts among younger and older adults. *Educational Gerontology*, 26(8), 737–749. <https://doi.org/10.1080/036012700300001395>

Liu, D., Wellman, H. M., Tardif, T., & Sabbagh, M. A. (2008). Theory of mind development in Chinese children: A meta-analysis of false-belief understanding across cultures and languages. *Developmental Psychology*, 44, 523–531. <https://doi.org/10.1037/0012-1649.44.2.523>

- Lüdecke D, Ben-Shachar M, Patil I, Waggoner P, Makowski D (2021). “performance: An R Package for Assessment, Comparison and Testing of Statistical Models.” *Journal of Open Science Software*, 6(60), 3139. <http://10.21105/joss.03139>
- Lugnegård, T., Unenge Hallerbäck, M., Hjärthag, F., & Gillberg, C. (2013). Social cognition impairments in Asperger syndrome and schizophrenia. *Schizophrenia Research*, 143(2-3), 277–284. <https://doi.org/10.1016/j.schres.2012.12.001>
- Massonnié, J., Bianco, M., Lima, L., & Bressoux, P. (2019). Longitudinal predictors of reading comprehension in French at first grade: Unpacking the oral comprehension component of the simple view. *Learning and Instruction*, 60, 166–179. <https://doi.org/10.1016/j.learninstruc.2018.01.005>
- Mar, R. A., & Oatley, K. (2008). The Function of Fiction is the Abstraction and Simulation of Social Experience. *Perspectives on Psychological Science*, 3(3), 173–192. <https://doi.org/10.1111/j.1745-6924.2008.00073.x>
- Mar, R. A., Tackett, J. L., & Moore, C. (2010). Exposure to media and theory-of-mind development in pre-schoolers. *Cognitive Development*, 25(1), 69–78. <https://doi.org/10.1016/j.cogdev.2009.11.002>
- Marjoram, D., Gardner, C., Burns, J., Miller, P., Lawrie, S., & Johnstone, E. (2005). Symptomatology and social inference: a mentalising study of schizophrenia and psychotic affective disorder. *Cognitive Neuropsychiatry*, 10(5), 347-359. <https://doi.org/10.1080/13546800444000092>
- Martin, I., & McDonald, S. (2003). Weak coherence, no mentalising, or executive dysfunction? Solving the puzzle of pragmatic language disorders. *Brain and Language*, 85(3), 451–466. [https://doi.org/10.1016/s0093-934x\(03\)00070-1](https://doi.org/10.1016/s0093-934x(03)00070-1)
- Martin, R. A., Puhlik-Doris, P., Larsen, G., Gray, J., & Weir, K. (2003). Individual differences in uses of humor and their relation to psychological well-being:

- Development of the Humor Styles Questionnaire. *Journal of Research in Personality*, 37, 48-75. [https://doi.org/10.1016/S0092-6566\(02\)00534-2](https://doi.org/10.1016/S0092-6566(02)00534-2).
- Martino, D. J., Strejilevich, S. A., Fassi, G., Marengo, E., & Igoa, A. (2011). Mentalising and facial emotion recognition in euthymic bipolar I and bipolar II disorders. *Psychiatry Research*, 189(3), 379–384.
<https://doi.org/10.1016/j.psychres.2011.04.033>
- Mason, R. A., Williams, D. L., Kana, R. K., Minshew, N., & Just, M. A. (2008). Mentalising disruption and recruitment of the right hemisphere during narrative comprehension in autism. *Neuropsychologia*, 46(1), 269-280.
<https://doi.org/10.1016/j.neuropsychologia.2007.07.018>
- Massaro, D., Valle, A., & Marchetti, A. (2013). Do social norms, false belief understanding, and metacognitive vocabulary influence irony comprehension? A study of five- and seven-year-old children. *European Journal of Developmental Psychology*, 11(3), 292–304. <https://doi.org/10.1080/17405629.2013.821407>
- Mayer, A., & Träuble, B. E. (2012). Synchrony in the onset of mental state understanding across cultures? A study among children in Samoa. *International Journal of Behavioral Development*, 37(1), 21–28. <https://doi.org/10.1177/0165025412454030>
- McDonald, S., Flanagan, S., Rollins, J., & Kinch, J. (2003). TASIT: A New Clinical Tool for Assessing Social Perception after traumatic brain injury. *Journal of Head Trauma Rehabilitation*, 18, 219-238. <https://doi.org/10.1037/0894-4105.18.3.572>
- McVay, J. C., & Kane, M. J. (2012). Why does working memory capacity predict variation in reading comprehension? On the influence of mind wandering and executive attention. *Journal of Experimental Psychology: General*, 141(2), 302–320.
<https://doi.org/10.1037/a0025250>

- Meinhardt-Injac, B., Daum, M. M., & Meinhardt, G. (2020). Mentalising development from adolescence to adulthood: Testing the two-component model. *The British Journal of Developmental Psychology*, 38(2), 289–303. <https://doi.org/10.1111/bjdp.12320>
- Meteyard, L., & Davies, R. A. (2020). Best practice guidance for linear mixed-effects models in psychological science. *Journal of Memory and Language*, 112, 104092. <https://doi.org/10.1016/j.jml.2020.104092>
- Metsala, J. L., & David, M. D. (2021). Improving English reading fluency and comprehension for children with reading fluency disabilities. *Dyslexia*. <https://doi.org/10.1002/dys.1695>
- Miguel, F. K., Caramanico, R. B., Huss, E. Y., & Zuanazzi, A. C. (2017). Validity of the reading the mind in the eyes test in a Brazilian sample1. *Paidéia (Ribeirão Preto)*, 27, 16-23. <https://doi.org/10.1590/1982-43272766201703>
- Miller, S. A. (2009). Children’s understanding of second-order mental states. *Psychological Bulletin*, 135(5), 749–773. <https://doi.org/10.1037/a0016854>
- Milligan, K., Astington, J. W., & Dack, L. A. (2007). Language and mentalising: metaanalysis of the relation between language ability and false-belief understanding. *Child Development*, 78(2), 622–646. <https://doi.org/10.1111/j.1467-8624.2007.01018.x>
- Mitchell, R. L., & Phillips, L. H. (2015). The overlapping relationship between emotion perception and mentalising. *Neuropsychologia*, 70, 1-10. <https://doi.org/10.1016/j.neuropsychologia.2015.02.018>
- Molenberghs, P., Johnson, H., Henry, J. D., & Mattingley, J. B. (2016). Understanding the minds of others: A neuroimaging meta-analysis. *Neuroscience & Biobehavioural Reviews*, 65, 276–291. <https://doi.org/10.1016/j.neubiorev.2016.03.020>

- Morgan, J. L. (1978). Two types of convention in indirect speech acts. In *Pragmatics* (pp. 261-280). Brill. https://doi.org/10.1163/9789004368873_010
- Muller, F., Simion, A., Reviriego, E., Galera, C., Mazaux, J.-M., Barat, M., & Joseph, P.-A. (2010). Exploring theory of mind after severe traumatic brain injury. *Cortex*, 46(9), 1088–1099. <https://doi.org/10.1016/j.cortex.2009.08.014>
- Murray, H.A. (1943). *Thematic Apperception Test*. Cambridge, MA: Harvard University Press. <https://psycnet.apa.org/record/1944-01759-000>
- Nation, K., Clarke, P., Marshall, C. M., & Durand, M. (2004). Hidden Language Impairments in Children. *Journal of Speech, Language, and Hearing Research*, 47(1), 199–211. [https://doi.org/10.1044/1092-4388\(2004/017\)](https://doi.org/10.1044/1092-4388(2004/017))
- Nation, K., & Norbury, C. F. (2005). Why reading comprehension fails: Insights from developmental disorders. *Topics in language disorders*, 25(1), 21-32. <https://doi.org/10.1097/00011363-200501000-00004>
- Norbury, C., & Nation, K. (2011). Understanding Variability in Reading Comprehension in Adolescents With Autism Spectrum Disorders: Interactions With Language Status and Decoding Skill. *Scientific Studies of Reading*, 15(3), 191–210. <https://doi.org/10.1080/10888431003623553>
- Nation, K., & Snowling, M. J. (1998). Semantic Processing and the Development of Word Recognition Skills: Evidence from Children with Reading Comprehension Difficulties. *Journal of Memory and Language*, 39(1), 85–101. <https://doi.org/10.1006/jmla.1998.2564>
- Nation, K., & Snowling, M. J. (2000). Factors influencing syntactic awareness skills in normal readers and poor comprehenders. *Applied Psycholinguistics*, 21(2), 229–241. <https://doi.org/10.1017/s0142716400002046>

- Navarra-Ventura, G., Fernandez-Gonzalo, S., Turon, M., Pousa, E., Palao, D., Cardoner, N., & Jodar, M. (2018). Gender differences in social cognition: a cross-sectional pilot study of recently diagnosed patients with schizophrenia and healthy subjects. *The Canadian Journal of Psychiatry*, 63(8), 538-546.
<https://doi.org/10.1177/0706743717746661>
- Nelson, K. (2005). *Language pathways into the community of minds*. W: JW Astington, JA Baird (red.). *Why Language Matters for Theory of Mind* (s. 26-49).
- Nentjes, L., Bernstein, D., Arntz, A., van Breukelen, G., & Slaats, M. (2015). Examining the influence of psychopathy, hostility biases, and automatic processing on criminal offenders' Mentalising. *International Journal of Law and Psychiatry*, 38, 92–99.
<https://doi.org/10.1016/j.ijlp.2015.01.012>
- Neth, D., & Martinez, A. M. (2009). Emotion perception in emotionless face images suggests a norm-based representation. *Journal of Vision*, 9(1), 5–5.
<https://doi.org/10.1167/9.1.5>
- New J. J., Schultz R. T., Wolf J., Niehaus J. L., Klin A., German T. C., Scholl B. J. (2010). The scope of social attention deficits in autism: Prioritized orienting to people and animals in static natural scenes. *Neuropsychologia*, 48, 51–59.
<https://doi.org/10.1016/j.neuropsychologia.2009.08.008>
- Nicholson, T., Williams, D. M., Grainger, C., Lind, S. E., & Carruthers, P. (2019). Relationships between implicit and explicit uncertainty monitoring and mindreading: Evidence from autism spectrum disorder. *Consciousness and Cognition*, 70, 11–24.
<https://doi.org/10.1016/j.concog.2019.01.013>
- Nilsen, E. S., & Duong, D. (2013). Depressive symptoms and perspective-taking within a communicative context. *Cognition & Emotion*, 27, 335–344.
<https://doi.org/10.1080/02699931.2012.708648>

- Oakhill, J. V., & Cain, K. (2012). The Precursors of Reading Ability in Young Readers: Evidence From a Four-Year Longitudinal Study. *Scientific Studies of Reading*, 16(2), 91–121. <https://doi.org/10.1080/10888438.2010.529219>
- Oakhill, J., & Yuill, N. (2013). Higher order factors in comprehension disability: Processes and remediation. In *Reading comprehension difficulties* (pp. 69-92). Routledge.
- O'Brien, M., Miner Weaver, J. M., Nelson, J. A., Calkins, S. D., Leerkes, E. M., & Marcovitch S. (2011). Longitudinal association between children's understanding of emotion and mentalising. *Cognition and Emotion*, 25, 1074–1086. <https://doi.org/10.1080/02699931.2010.518417>
- OECD. (2019). *Survey of Adult Skills (PIAAC) - PIAAC, the OECD's programme of assessment and analysis of adult skills*. Oecd.org. <https://www.oecd.org/skills/piaac/>
- Oesch, N., & Dunbar, R. I. M. (2017). The emergence of recursion in human language: Mentalising predicts recursive syntax task performance. *Journal of Neurolinguistics*, 43, 95–106. <https://doi.org/10.1016/j.jneuroling.2016.09.008>
- O'Hare, A. E., Bremner, L., Nash, M., Happé, F., & Pettigrew, L. M. (2009). A clinical assessment tool for advanced mentalising performance in 5- to 12-year-olds. *Journal of autism and developmental disorders*, 39, 916-928. <https://doi.org/10.1007/s10803-009-0699-2>
- Olderbak, S., Wilhelm, O., Olaru, G., Geiger, M., Brenneman, M. W., & Roberts, R. D. (2015). A psychometric analysis of the reading the mind in the eyes test: toward a brief form for research and applied settings. *Frontiers in psychology*, 6, 1503. <https://doi.org/10.3389/fpsyg.2015.01503>
- Oldershaw, A., Hambrook, D., Rimes, K. A., Tchanturia, K., Treasure, J., Richards, S., Schmidt, U., & Chalder, T. (2011). Emotion recognition and emotional mentalising

- in chronic fatigue syndrome. *Psychology & Health*, 26(8), 989–1005.
<https://doi.org/10.1080/08870446.2010.519769>
- O'Neill, D. K. (2012). “Components of pragmatic ability and children's pragmatic language development,” in *Handbooks of Pragmatics 4: Cognitive Pragmatics*, eds W. Bublitz, A. H. Jucker, and K. P. Schneider (Series eds.) and H.-J. Schmid (Volume ed.) (Berlin: Mouton de Gruyter), 261–288. <https://doi.org/10.1515/9783110214215>
- O'Connor, I. M., & Klein, P. D. (2004). Exploration of Strategies for Facilitating the Reading Comprehension of High-Functioning Students with Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders*, 34(2), 115–127.
<https://doi.org/10.1023/b:jadd.0000022603.44077.6b>
- Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science*, 308(5719), 255–258. <https://doi.org/10.1126/science.1107621>
- Orellana, G., & Slachevsky, A. (2013). Executive functioning in schizophrenia. *Frontiers in psychiatry*, 4, 35. <https://doi.org/10.3389/fpsyt.2013.00035>
- Osterhaus, C., Koerber, S., & Sodian, B. (2016). Scaling of advanced theory-of-mind tasks. *Child development*, 87(6), 1971–1991. <https://doi.org/10.1111/cdev.12566>
- Osterhaus, C., & Bosacki, S. L. (2022). Looking for the lighthouse: A systematic review of advanced theory-of-mind tests beyond preschool. *Developmental Review: DR*, 64, 101021. <https://doi.org/10.1016/j.dr.2022.101021>
- Ouellette, G. P. (2006). What's meaning got to do with it: The role of vocabulary in word reading and reading comprehension. *Journal of Educational Psychology*, 98(3), 554–566. <https://doi.org/10.1037/0022-0663.98.3.554>
- Ouellet, J., Scherzer, P. B., Rouleau, I., Métras, P., Bertrand-Gauvin, C., Djerroud, N., et al. (2010). Assessment of social cognition in patients with multiple sclerosis. *J. Int. Neuropsychol. Soc.* 16, 287–296. <https://doi.org/10.1017/S1355617709991329>

- Palmer, S. J., Fanucci-Kiss, A., Kipervassar, E., Jalnapurkar, I., Hodge, S. M., Frazier, J. A., & Cochran, D. (2023). Effect of Emotional Valence on Emotion Recognition in Adolescents with Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*, 1-13. <https://doi.org/10.1007/s10803-022-05831-5>
- Park, H. (2008). Home literacy environments and children's reading performance: A comparative study of 25 countries. *Educational Research and Evaluation: An International Journal on Theory and Practice*, 14(6), 489–505. <https://doi.org/10.1080/13803610802576734>
- Payne, B. R., Gao, X., Noh, S. R., Anderson, C. J., & Stine-Morrow, E. A. (2012). The effects of print exposure on sentence processing and memory in older adults: Evidence for efficiency and reserve. *Aging, Neuropsychology, and Cognition*, 19(1-2), 122-149. <https://doi.org/10.1080/13825585.2011.628376>
- Parola, A., Salvini, R., Gabbatore, I., Colle, L., Berardinelli, L., & Bosco, F. M. (2020). Pragmatics, Mentalising and executive functions in schizophrenia: Disentangling the puzzle using machine learning. *PLOS ONE*, 15(3), e0229603. <https://doi.org/10.1371/journal.pone.0229603>
- Peñuelas-Calvo, I., Sareen, A., Sevilla-Llewellyn-Jones, J., & Fernández-Berrocal, P. (2019). The “Reading the Mind in the Eyes” test in autism-spectrum disorders comparison with healthy controls: a systematic review and meta-analysis. *Journal of Autism and Developmental Disorders*, 49(3), 1048-1061. <https://doi.org/10.1007/s10803-018-3814-4>
- Perfetti, C., McKeown, M. G., & Kucan, L. (2010). Decoding, vocabulary, and comprehension. *Bringing reading research to life*, 291-303.

- Perfetti C. A., Yang C.-L., & Schmalhofer F. (2008). Comprehension skill and word-to-text integration processes. *Applied Cognitive Psychology*, 22, 303-318. <http://10.1002/acp.1419>
- Perner, J., & Ruffman, T. (2005). Infants' insight into the mind: How deep? *Science*, 308(5719), 214-216. <https://doi.org/10.1126/science.1111656>
- Pershey, M. G. (1998). Teaching children to identify and respond to pragmatic language functions in narrative text. *Reading Improvement*, 35(4), 146. <https://doi.org/10.1080/09658410008667147>
- Peterson, E., & Miller, S. F. (2012). The Eyes Test as a Measure of Individual Differences: How much of the Variance Reflects Verbal IQ? *Frontiers in Psychology*, 3. <https://doi.org/10.3389/fpsyg.2012.00220>
- Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and brain sciences*, 27(2), 169-190. <https://doi.org/10.1017/S0140525X04000056>
- Polyanskaya, I., Eigsti, I.-M., Brauner, T., & Blackburn, P. (2021). Second-Order False Beliefs and Linguistic Recursion in Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*, 1-16. <https://doi.org/10.1007/s10803-021-05277-1>
- Poulin-Dubois, D., Rakoczy, H., Burnside, K., Crivello, C., Dörrenberg, S., Edwards, K., ... & Ruffman, T. (2018). Do infants understand false beliefs? We don't know yet—A commentary on Baillargeon, Buttelmann and Southgate's commentary. *Cognitive Development*, 48, 302-315. <https://doi.org/10.1111/j.1468-5922.2007.00685.x>
- Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind?. *Behavioural and brain sciences*, 1(4), 515-526. <https://doi.org/10.1017/S0140525X00076512>

- Prinz, A., Golke, S., & Wittwer, J. (2020). To What Extent Do Situation-Model-Approach Interventions Improve Relative Metacomprehension Accuracy? Meta-Analytic Insights. *Educational Psychology Review*, 32(4), 917–949.
<https://doi.org/10.1007/s10648-020-09558-6>
- Quesque, F., Coutrot, A., Cox, S., de Souza, L. C., Baez, S., Cardona, J. F., Mulet-Perreault, H., Flanagan, E., Neely-Prado, A., Clarens, M. F., Cassimiro, L., Musa, G., Kemp, J., Botzung, A., Philippi, N., Cosseddu, M., Trujillo-Llano, C., Grisales-Cardenas, J. S., Fittipaldi, S., ... Bertoux, M. (2022). Does culture shape our understanding of others' thoughts and emotions? An investigation across 12 countries. *Neuropsychology*, 36(7), 664–682. <https://doi.org/10.1037/neu0000817>
- Quesque, F., & Rossetti, Y. (2020). What Do Theory-of-Mind Tasks Actually Measure? Theory and Practice. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, 15(2), 384–396
<https://doi.org/10.1177/1745691619896607>
- Rahman, F., Kessler, K., Apperly, I. A., Hansen, P. C., Javed, S., Holland, C. A., & Hartwright, C. E. (2021). Sources of Cognitive Conflict and Their Relevance to Theory-of-Mind Proficiency in Healthy Aging: A Preregistered Study. *Psychological Science*, 32(12), 1918–1936. <https://doi.org/10.1177/09567976211017870>
- Rakoczy, H. (2022). Foundations of theory of mind and its development in early childhood. *Nature Reviews Psychology*, 1. <https://doi.org/10.1038/s44159-022-00037-z>
- Rambeau, S., Del Goleto, S., Pignon, B., Lajnef, M., Petrucci, J., Szöke, A., Fond, G., Lançon, C., Dorey, J.-M., Rey, R., Garbisson, A., Capdevielle, D., Leignier, S., Dubreucq, J., Mallet, J., Dubertret, C., Urbach, M., Brunet-Gouet, E., Aouizerate, B., ... Schürhoff, F. (2022). Relationship between neurocognition and mentalising as a

- function of symptomatic profile in schizophrenia: results from the national FACE-SZ cohort. *Cognitive Neuropsychiatry*, 27(1), 49–68.
<https://doi.org/10.1080/13546805.2021.2011184>
- Ramscar, M., Hendrix, P., Love, B., & Baayen, R. H. (2013). Learning is not decline: The mental lexicon as a window into cognition across the lifespan. *The Mental Lexicon*, 8(3), 450-481. <https://doi.org/10.1075/ml.8.3.08ram>
- RAND Reading Study Group. (2002). *Reading for understanding: Towards an R&D program in reading comprehension*. Retrieved May 28, 2018 from
https://www.rand.org/content/dam/rand/pubs/monograph_reports/2005/MR1465.pdf
- Rasmussen, C. E., & Jiang, Y. V. (2019). Judging social interaction in the Heider and Simmel movie. *Quarterly Journal of Experimental Psychology*, 72(9), 2350-2361.
<https://doi.org/10.1177/1747021819838764>
- Rayner, K., Kambe, G., & Duffy, S. A. (2000). The effect of clause wrap-up on eye movements during reading. *The Quarterly Journal of Experimental Psychology: Section A*, 53(4), 1061-1080. <https://doi.org/10.1080/713755934>
- Read On. Get On. (2014), *How reading can help children escape poverty*. Published by Save the Children on behalf of the Read On. Get On. campaign, p.17. Available at:
https://www.savethechildren.org.uk/sites/default/files/images/Read_On_Get_On.pdf
- Recanatì, F. (2001). What is said. *Synthese*, 128 (1/2), 75-91.
<https://www.jstor.org/stable/20117147>
- Reddy, V. (2001). Infant clowns: The interpersonal creation of humor in infancy. *Childhood*, 53 (3), 247-256. <https://doi.org/10.3917/enf.533.0247>
- Rice, K., & Redcay, E. (2015). Spontaneous mentalizing captures variability in the cortical thickness of social brain regions. *Social Cognitive and Affective Neuroscience*, 10(3), 327–334. <https://doi.org/10.1093/scan/nsu081>

- Richter, T., Appel, M., & Calio, F. (2014). Stories can influence the self-concept. *Social Influence*, 9(3), 172–188. <https://doi.org/10.1080/15534510.2013.799099>
- Rigby, J., Conroy, S., Miele-Norton, M., Pawlby, S., & Happé, F. (2016). Mentalising as a predictor of maternal sensitivity in women with severe mental illness. *Psychological medicine*, 46(9), 1853–1863. <https://doi.org/10.1017/S0033291716000337>
- Rokita, K. I., Dauvermann, M. R., Mothersill, D., Holleran, L., Holland, J., Costello, L., Cullen, C., Kane, R., McKernan, D., Morris, D. W., Kelly, J., Gill, M., Corvin, A., Hallahan, B., McDonald, C., & Donohoe, G. (2021). Childhood trauma, parental bonding, and social cognition in patients with schizophrenia and healthy adults. *Journal of Clinical Psychology*, 77(1), 241–253. <https://doi.org/10.1002/jclp.23023>
- Rose, J. (2006). *Independent review of the teaching of early reading*. Nottingham: DfES Publications.
- Rosenblau, G., Kliemann, D., Heekeren, H. R., & Dziobek, I. (2015). Approximating implicit and explicit mentalizing with two naturalistic video-based tasks in typical development and autism spectrum disorder. *Journal of Autism and Developmental Disorders*, 45(4), 953–965. <https://doi.org/10.1007/s10803-014-2249-9>
- Rossell, S. L., & Van Rheenen, T. E. (2013). Mentalising performance using a story comprehension task in bipolar mania compared to schizophrenia and healthy controls. *Cognitive Neuropsychiatry*, 18(5), 409–421. <https://doi.org/10.1080/13546805.2012.725820>
- Rowley, D. A., Rogish, M., Alexander, T., & Riggs, K. J. (2017). Cognitive correlates of pragmatic language comprehension in adult traumatic brain injury: A systematic review and meta-analyses. *Brain injury*, 31(12), 1564–1574. <https://doi.org/10.1080/02699052.2017.1341645>

- Royeras, J. T., & Sumayo, G. S. (2024). Vocabulary Knowledge and Inferential Reading Comprehension of Senior High School Students: A Descriptive-Correlational Inquiry. *East Asian Journal of Multidisciplinary Research*, 3(3), 1143–1154. <https://doi.org/10.55927/eajmr.v3i3.8164>
- Rubio-Fernández, P. (2017). The director task: A test of Theory-of-Mind use or selective attention? *Psychonomic Bulletin & Review*, 24(4), 1121–1128. <https://doi.org/10.3758/s13423-016-1190-7>
- Ruytenbeek, N., Ostashchenko, E., & Kissine, M. (2017). Indirect request processing, sentence types and illocutionary forces. *Journal of pragmatics*, 119, 46-62. <https://doi.org/10.1016/j.pragma.2017.07.011>
- Ryskin, R. A., Benjamin, A. S., Tullis, J., & Brown-Schmidt, S. (2015). Perspective-taking in comprehension, production, and memory: An individual differences approach. *Journal of Experimental Psychology: General*, 144(5), 898. <https://doi.org/10.1037/xge0000093>
- Samson, D., Apperly, I. A., Braithwaite, J. J., Andrews, B. J., & Bodley Scott, S. E. (2010). Seeing it their way: Evidence for rapid and involuntary computation of what other people see. *Journal of Experimental Psychology. Human Perception and Performance*, 36(5), 1255–1266. [1255–1266. https://doi.org/10.1037/a0018729](https://doi.org/10.1037/a0018729)
- Savage, R. (2006). Reading Comprehension Is Not Always the Product of Nonsense Word Decoding and Linguistic Comprehension: Evidence from Teenagers Who Are Extremely Poor Readers. *Scientific Studies of Reading*, 10(2), 143–164. https://doi.org/10.1207/s1532799xssr1002_2
- Saxe, R. (2006). Uniquely human social cognition. *Current Opinion in Neurobiology*, 16(2), 235–239. <https://doi.org/10.1016/j.conb.2006.03.001>

- Schaafsma, S. M., Pfaff, D. W., Spunt, R. P., & Adolphs, R. (2015). Deconstructing and reconstructing mentalising. *Trends in Cognitive Sciences*, 19(2), 65–72.
<https://doi.org/10.1016/j.tics.2014.11.007>
- Schallert, D. L., Kleiman, G. M., & Rubin, A. D. (1977). Analyses of differences between written and oral language. *Center for the Study of Reading Technical Report; no. 029*.
- Scherzer, P., Achim, A., Léveillé, E., Boisseau, E., & Stip, E. (2015). Evidence from paranoid schizophrenia for more than one component of mentalising. *Frontiers in Psychology*, 6, 1643. <https://doi.org/10.3389/fpsyg.2015.01643>
- Schick, B., De Villiers, P., De Villiers, J., & Hoffmeister, R. (2007). Language and mentalising: A study of deaf children. *Child development*, 78(2), 376-396.
<https://doi.org/10.1111/j.1467-8624.2007.01004.x>
- Schielzeth, H., Dingemanse, N. J., Nakagawa, S., Westneat, D. F., Alagüe, H., Teplitsky, C., ... & Araya-Ajoy, Y. G. (2020). Robustness of linear mixed-effects models to violations of distributional assumptions. *Methods in Ecology and Evolution*, 11(9), 1141-1152. <https://doi.org/10.1111/2041-210x.13434>
- Scholl, B. J., & Gao, T. (2013). Perceiving animacy and intentionality: Visual processing or higher-level judgment. *Social perception: Detection and interpretation of animacy, agency, and intention*, 4629, 197-229.
- Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize? *Journal of Research in Personality*, 47(5), 609–612.
<https://doi.org/10.1016/j.jrp.2013.05.009>
- Schuwerk, T., Vuori, M., & Sodian, B. (2015). Implicit and explicit Mentalising reasoning in autism spectrum disorders: the impact of experience. *Autism: The International*

Journal of Research and Practice, 19(4), 459–468.

<https://doi.org/10.1177/1362361314526004>

Schurz, M., Radua, J., Aichhorn, M., Richlan, F., & Perner, J. (2014). Fractionating mentalising: a meta-analysis of functional brain imaging studies. *Neuroscience & Biobehavioral Reviews*, 42, <https://doi.org/10.1016/j.neubiorev.2014.01.009>

Schwartz Offek, E., & Segal, O. (2022). Comparing theory of mind development in children with autism spectrum disorder, developmental language disorder, and typical development. *Neuropsychiatric disease and treatment*, 2349-2359.

<https://doi.org/10.2147/NDT.S331988>

Seo, E., Koo, S. J., Kim, Y. J., Min, J. E., Park, H. Y., Bang, M., ... & An, S. K. (2020). Reading the mind in the eyes test: Relationship with neurocognition and facial emotion recognition in non-clinical youths. *Psychiatry Investigation*, 17(8), 835.

<https://doi.org/10.30773/pi.2019.0281>

Shany-Ur, T., Poorzand, P., Grossman, S. N., Growdon, M. E., Jang, J. Y., Ketelle, R. S., Miller, B. L., & Rankin, K. P. (2012). Comprehension of insincere communication in neurodegenerative disease: lies, sarcasm, and mentalising. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, 48(10), 1329–1341.

<https://doi.org/10.1016/j.cortex.2011.08.003>

Shao, Z., Janse, E., Visser, K., & Meyer, A. S. (2014). What do verbal fluency tasks measure? Predictors of verbal fluency performance in older adults. *Frontiers in psychology*, 5, 772. <https://doi.org/10.3389/fpsyg.2014.00772>

Shapiro, A. M., & Murphy, G. L. (1993). Can You Answer a Question for Me? Processing Indirect Speech Acts. *Journal of Memory and Language*, 32(2), 211–229.

<https://doi.org/10.1006/jmla.1993.1012>

- Shaw, P., Lawrence, E. J., Radbthene, C., Bramham, J., Polkey, C. E., & David, A. S. (2004). The impact of early and late damage to the human amygdala on “mentalising” reasoning. *Brain: A Journal of Neurology*, 127(7), 1535–1548.
<https://doi.org/10.1093/brain/awh168>
- Silagi, M. L., Romero, V. U., Mansur, L. L., & Radanovic, M. (2014). Inference comprehension during reading: influence of age and education in normal adults. *CoDAS*, 26(5), 407–414. <https://doi.org/10.1590/2317-1782/20142013058>
- Slade, L., & Ruffman, T. (2005). How language does (and does not) relate to theory of mind: A longitudinal study of syntax, semantics, working memory and false belief. *British Journal of Developmental Psychology*, 23(1), 117–141.
<https://doi.org/10.1348/026151004x21332>
- Snowling M. J., & Hulme C. (2005). *Learning to read with a language impairment*. The science of reading: a handbook (eds , Snowling M. J., & Hulme C.), pp. 397–412. Oxford, UK: Blackwell.
- Snowling, M. J., Stothard, S. E., Clarke, P., Bowyer-Crane, C., Harrington, A., Truelove, E., Nation, K. & Hulme, C. (2011). *York Assessment of Reading for Comprehension: Early Reading and Passage Reading Primary Second Edition*. London, United Kingdom: GL Assessment.
- Spotorno, N., McMillan, C. T., Rascovsky, K., Irwin, D. J., Clark, R., & Grossman, M. (2015). Beyond words: Pragmatic inference in behavioral variant of frontotemporal degeneration. *Neuropsychologia*, 75, 556-564.
<https://doi.org/10.1016/j.neuropsychologia.2015.07.002>
- Sperber, D., & Wilson, D. (1995). *Relevance: Communication and cognition*, 2nd ed.
- Sperber, D., & Wilson, D. (2002). Pragmatics, modularity and mind-reading. *Mind & language*, 17(1-2), 3-23. <https://doi.org/10.1111/1468-0017.00186>

- Spenser, K., Bull, R., Betts, L., & Winder, B. (2020). Underpinning prosociality: Age related performance in mentalising, empathic understanding, and moral reasoning. *Cognitive Development*, 56, 100928. <https://doi.org/10.1016/j.cogdev.2020.100928>
- Stanley, J. T., Zhang, X., Fung, H. H., & Isaacowitz, D. M. (2013). Cultural Differences in Gaze and Emotion Recognition: Americans Contrast More than Chinese. *Emotion (Washington, D.C.)*, 13(1), 36–46. <https://doi.org/10.1037/a0029209>
- Steen-Baker, A. A., Ng, S., Payne, B. R., Anderson, C. J., Federmeier, K. D., & StineMorrow, E. A. (2017). The effects of context on processing words during sentence reading among adults varying in age and literacy skill. *Psychology and Aging*, 32(5), 460. <https://doi.org/10.1037/pag0000184>
- Stewart, G. R., Wallace, G. L., Cottam, M., & Charlton, R. A. (2019). Theory of mind performance in younger and older adults with elevated autistic traits. *Autism Research*. <https://doi.org/10.1002/aur.2206>
- Stine-Morrow, E. A., Shake, M. C., Miles, J. R., Lee, K., Gao, X., & McConkie, G. (2010). Pay now or pay later: Aging and the role of boundary salience in self-regulation of conceptual integration in sentence processing. *Psychology and Aging*, 25, 168 –176. <http://dx.doi.org/10.1037/a0018127>
- Storch, S. A., & Whitehurst, G. J. (2002). Oral language and code-related precursors to reading: Evidence from a longitudinal structural model. *Developmental Psychology*, 38(6), 934–947. <https://doi.org/10.1037/0012-1649.38.6.934>
- Stone, V. E., Baron-Cohen, S., & Knight, R. T. (1998). Frontal lobe contributions to mentalising. *Journal of Cognitive Neuroscience*, 10(5), 640–656. <https://doi.org/10.1162/089892998562942>
- Stonnington, C. M., Locke, D. E. C., Hsu, C.-H., Ritenbaugh, C., & Lane, R. D. (2013). Somatization is associated with deficits in affective Mentalising. *Journal of*

Psychosomatic Research, 74(6), 479–485.

<https://doi.org/10.1016/j.jpsychores.2013.04.004>

Strasser, K., & Río, F. del. (2013). The Role of Comprehension Monitoring, Mentalising, and Vocabulary Depth in Predicting Story Comprehension and Recall of Kindergarten Children. *Reading Research Quarterly*, 49(2), 169–187.

<https://doi.org/10.1002/rrq.68>

Strous, R. D., Koppel, M., Fine, J., Nachliel, S., Shaked, G., & Zivotofsky, A. Z. (2009). Automated characterization and identification of schizophrenia in writing. *The Journal of nervous and mental disease*, 197(8), 585–588.

<https://doi.org/10.1097/nmd.0b013e3181b09068>

Suls, J. (1977). Cognitive and Disparagement Theories of Humour: A Theoretical and Empirical Synthesis. *It's a Funny Thing, Humour*, 41–45.

<https://doi.org/10.1016/b978-0-08-021376-7.50012-7>

Suls, J.M., (1983). *Cognitive processes in humor appreciation*. In: McGhee, P.E., Goldstein, J.H. (Eds.), *Handbook of Humor Research*, Vol. 1: Basic Issues. Springer-Verlag, New York, pp. 39–57. <https://doi.org/10.12691/education-2-7-12>

Surtees, A., Samson, D., & Apperly, I. (2016). Unintentional perspective-taking calculates whether something is seen, but not how it is seen. *Cognition*, 148, 97–105.

<https://doi.org/10.1016/j.cognition.2015.12.010>

Symeonidou, I., Dumontheil, I., Chow, W.-Y., & Breheny, R. (2016). Development of online use of mentalising during adolescence: An eye-tracking study. *Journal of Experimental Child Psychology*, 149, 81–97.

<https://doi.org/10.1016/j.jecp.2015.11.007>

- Szücs, M., & Babarczy, A. (2017). 8. The role of Mentalising, grammatical competence and metapragmatic awareness in irony comprehension. *De Gruyter EBooks*, 129–148.
<https://doi.org/10.1515/9781501505089-008>
- Tager-Flusberg, H. (1993). What language reveals about the understanding of minds in children with autism. In S. Baron-Cohen, H. Tager-Flusberg, & D.J. Cohen (Eds.), *Understanding other minds: Perspectives from autism*. Oxford: Oxford University Press.
- Tannen, D., & Wallat, C. (1987). Interactive frames and knowledge schemas in interaction: Examples from a medical examination/interview. *Social psychology quarterly*, 205–216. <https://doi.org/10.2307/2786752>
- Tannenbaum, K. R., Torgesen, J. K., & Wagner, R. K. (2006). Relationships Between Word Knowledge and Reading Comprehension in Third-Grade Children. *Scientific Studies of Reading*, 10(4), 381–398. https://doi.org/10.1207/s1532799xssr1004_3
- Taylor, C. (2016). Mock politeness and culture: Perceptions and practice in UK and Italian data. *Intercultural Pragmatics*, 13(4), 463–498. <https://doi.org/10.1515/ip-2016-0021>
- Thiébaud, F. I., White, S. J., Walsh, A., Klargaard, S. K., Wu, H. C., Rees, G., & Burgess, P. W. (2016). Does faux pas detection in adult autism reflect differences in social cognition or decision-making abilities?. *Journal of autism and developmental disorders*, 46, 103–112. <https://doi.org/10.1007/s10803-015-2551-1>
- Tobia, V., & Bonifacci, P. (2015). The simple view of reading in a transparent orthography: the stronger role of oral comprehension. *Reading and Writing*, 28(7), 939–957.
<https://doi.org/10.1007/s11145-015-9556-1>
- Tiv, M., O'Regan, E., & Titone, D. (2021). In a bilingual state of mind: Investigating the continuous relationship between bilingual language experience and mentalizing.

Bilingualism: Language and Cognition, 1–14.

<https://doi.org/10.1017/s1366728921000225>

- Tromp, J., Hagoort, P., & Meyer, A. S. (2016). Pupillometry reveals increased pupil size during indirect request comprehension. *Quarterly Journal of Experimental Psychology*, 69(6), 1093–1108. , <http://dx.doi.org/10.1080/17470218.2015.1065282>
- Trott, S., & Bergen, B. (2019). Individual Differences in Mentalizing Capacity Predict Indirect Request Comprehension. *Discourse Processes*, 56(8), 675–707. <https://doi.org/10.1080/0163853x.2018.1548219>
- Trott, S., & Bergen, B. (2020). When Do Comprehenders Mentalize for Pragmatic Inference? *Discourse Processes*, 57(10), 900–920. <https://doi.org/10.1080/0163853x.2020.1822709>
- Turkstra, L. S. (2008). Conversation-based assessment of social cognition in adults with traumatic brain injury. *Brain Injury: [BI]*, 22(5), 397–409. <https://doi.org/10.1080/02699050802027059>
- Turner, R., & Felisberti, F. M. (2017). Measuring Mindreading: A Review of Behavioral Approaches to Testing Cognitive and Affective Mental State Attribution in Neurologically Typical Adults. *Frontiers in Psychology*, 8, 47. <https://doi.org/10.3389/fpsyg.2017.00047>
- Van Ackeren, M. J., Casasanto, D., Bekkering, H., Hagoort, P., & Rueschemeyer, S. A. (2012). Pragmatics in action: indirect requests engage mentalising areas and the cortical motor network. *Journal of Cognitive Neuroscience*, 24(11), [2237–2247](https://doi.org/10.1162/jocn_a_00274). https://doi.org/10.1162/jocn_a_00274
- van Ackeren, M. J., Smaragdi, A., & Rueschemeyer, S. A. (2016). Neuronal interactions between mentalising and action systems during indirect request processing. *Social*

cognitive and affective neuroscience, 11(9), 1402-1410.

<https://doi.org/10.1093/scan/nsw062>

van Bergen, E., Vasalampi, K., & Torppa, M. (2020). How Are Practice and Performance Related? Development of Reading from Age 5 to 15. *Reading Research Quarterly*.

<https://doi.org/10.1002/rrq.309>

van Wingerden, E., Segers, E., van Balkom, H., & Verhoeven, L. (2018). Cognitive Constraints on the Simple View of Reading: A Longitudinal Study in Children with Intellectual Disabilities. *Scientific Studies of Reading*, 22(4), 321–334.

<https://doi.org/10.1080/10888438.2018.1446435>

Vellante, M., Baron-Cohen, S., Melis, M., Marrone, M., Petretto, D. R., Masala, C., et al.

(2012). The “Reading the Mind in the Eyes” test: systematic review of psychometric properties and a validation study in Italy. *Cognitive Neuropsychiatry* 18, 1–12.

<https://doi.org/10.1080/13546805.2012.721728>

Visser, C. T. W., Tomas, E., & Law, J. (2020). The emergence of inner speech and its measurement in atypically developing children. *Frontiers in Psychology*, 11, 279.

<https://doi.org/10.3389/fpsyg.2020.00279>

Walsh-Messinger, J., Stepanek, C., Wiedemann, J., Goetz, D., Goetz, R. R., & Malaspina,

D. (2019). Normal sexual dimorphism in mentalising circuitry is reversed in Schizophrenia. *Social Neuroscience*, 14(5), 583–593.

<https://doi.org/10.1080/17470919.2018.1536613>

Walczyk, J. J., Marsiglia, C. S., Johns, A. K., & Bryan, K. S. (2004). Children’s

Compensations for Poorly Automated Reading Skills. *Discourse Processes*, 37(1), 47–66. https://doi.org/10.1207/s15326950dp3701_3

Wang, J. J., Ali, M., Frisson, S., & Apperly, I. A. (2016). Language complexity modulates 8- and 10-year-olds’ success at using their mentalising abilities in a communication

- task. *Journal of Experimental Child Psychology*, 149, 62–71.
<https://doi.org/10.1016/j.jecp.2015.09.006>
- Wang, M., Gamo, N. J., Yang, Y., Jin, L. E., Wang, X. J., Laubach, M., ... & Arnsten, A. F. (2011). Neuronal basis of age-related working memory decline. *Nature*, 476(7359), 210-213. <https://doi.org/10.1038/nature10243>
- Wang, Y., & Su, Y. (2006). Mentalising in old adults: The performance on Happé's Stories and Faux Pas Stories. *Psychologia-an International Journal of Psychology in the Orient*, 49(4), 228–237. <https://doi.org/10.2117/psysoc.2006.228>
- Wardlow, L. (2013). Individual differences in speakers' perspective taking: The roles of executive control and working memory. *Psychonomic bulletin & review*, 20(4), 766-772. <https://doi.org/10.3758/s13423-013-0396-1>
- Warnell, K. R., & Redcay, E. (2019). Minimal coherence among varied mentalising measures in childhood and adulthood. *Cognition*, 191, 103997.
<https://doi.org/10.1080/17470919.2018.1536613>
- Wechsler, D. (1997). *WAIS-III: Administration and scoring manual: Wechsler adult intelligence scale*. Psychological Corporation. https://doi.org/10.1007/978-1-4615-1185-4_2
- Wechsler, D. (2001). *Wechsler Test of Adult Reading: WTAR*. Psychological Corporation.
- Wellman, H. M., & Bartsch, K. (1988) Young children's reasoning about beliefs. *Cognition*, 30(3), 239-277. [https://doi.org/10.1016/0010-0277\(88\)90021-2](https://doi.org/10.1016/0010-0277(88)90021-2)
- Wechsler, D. (2009). *Wechsler Individual Achievement Test* (3rd ed.). San Antonio, TX: Pearson.
- Weckerly J., Wulfeck B., Reilly J. (2001). Verbal fluency deficits in children with specific language impairment: slow rapid naming or slow to name? *Child Neuropsychology* 7, 142–152. <https://doi.org/10.1076/chin.7.3.142.8741>

- Wellman, H. M. (2018). Theory of mind across the lifespan? *Zeitschrift für Psychologie*, 226, 136–138. <https://doi.org/10.1027/2151-2604/a000330>
- Wellman, H., Cross, D., & Watson, J. (2001). Metaanalysis of mentalising development: The truth about false-belief. *Child Development*, 72(3), 655–684. <https://doi.org/10.1111/1467-8624.00304>
- Wellman, H. M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child development*, 75(2), 523-541. <https://doi.org/10.1111/j.1467-8624.2004.00691.x>
- Wertsch, J. V., & Stone, C. A. (1985). The concept of internalization in Vygotsky's account of the genesis of higher mental functions. *Culture, communication, and cognition: Vygotskian perspectives*, 162-179.
- White, L. J., Alexander, A., & Greenfield, D. B. (2017). The relationship between executive functioning and language: Examining vocabulary, syntax, and language learning in preschoolers attending Head Start. *Journal of Experimental Child Psychology*, 164, 16-31. <https://doi.org/10.1016/j.jecp.2017.06.010>
- White, S., Hill, E., Happé, F., & Frith, U. (2009). Revisiting the strange stories: revealing mentalizing impairments in autism. *Child Development*, 80(4), 1097–1117. <https://doi.org/10.1111/j.1467-8624.2009.01319.x>
- White, S. J., Coniston, D., Rogers, R., & Frith, U. (2011). Developing the Frith-Happé animations: A quick and objective test of Mentalising for adults with autism. *Autism Research*, 4(2), 149–154. <https://doi.org/10.1002/aur.174>
- Whitton, A. E., & Henry, J. D. (2012). The relationship between sub-clinical obsessive compulsive symptoms and social cognition in chronic schizophrenia. *British Journal of Clinical Psychology*, 52(2), 115–128. <https://doi.org/10.1111/bjc.12004>

- Wick, F. A., Garg, S., Soce, A., & Wolfe, J. (2016). Perception of dynamic scenes: What is your Heider capacity? *Journal of Vision*, 16(12), 573-573.
<https://doi.org/10.1037/xge0000557>
- Wilson, D. (2009). Irony and metarepresentation (Vol. 21, pp. 183-226). UCL Working Papers in Linguistics.
- Wilson, J., Andrews, G., Hogan, C., Wang, S., & Shum, D. H. K. (2018). Executive function in middle childhood and the relationship with mentalising. *Developmental Neuropsychology*, 43(3), 163–182. <https://doi.org/10.1080/87565641.2018.1440296>
- Wimmer, H., & Goswami, U. (1994). The influence of orthographic consistency on reading development: word recognition in English and German children. *Cognition*, 51(1), 91–103. [https://doi.org/10.1016/0010-0277\(94\)90010-8](https://doi.org/10.1016/0010-0277(94)90010-8)
- Winner, E. Brownell, H. Happé, F. Blum, A. Pincus D. (1998). Distinguishing Lies from Jokes: Mentalising Deficits and Discourse Interpretation in Right Hemisphere Brain-Damaged Patients. *Brain and Language*, 62(1), 89–106.
<https://doi.org/10.1006/brln.1997.1889>
- Woods, M. L., & Moe, A. J. (2011). Analytical reading inventory (9th ed.). Boston, MA: Allyn & Bacon.
- Wolkenstein, L., Schönenberg, M., Schirm, E., & Hautzinger, M. (2011). I can see what you feel, but I can't deal with it: Impaired mentalising in depression. *Journal of Affective Disorders*, 132(1-2), 104–111. <https://doi.org/10.1016/j.jad.2011.02.010>
- Yildirim, E., Soncu Büyükişcan, E., & Gürvit, H. (2020). Affective mentalising in human aging: Is there any relation with executive functioning? *Neuropsychology, Development, and Cognition. Section B, Aging, Neuropsychology and Cognition*, 27(2), 207–219. <https://doi.org/https://doi.org/10.1080/13825585.2019.1602706>

- Yong, M. H., Lawrie, L., Schaefer, A., & Phillips, L. H. (2022). The Effects of Adult Aging and Culture on Theory of Mind. *The Journals of Gerontology: Series B*, 77(2), 332–340. <https://doi.org/10.1093/geronb/gbab093>
- Yus, F. (2003). Humor and the search for relevance. *Journal of pragmatics*, 35(9), 1295–1331. [https://doi.org/10.1016/S0378-2166\(02\)00179-0](https://doi.org/10.1016/S0378-2166(02)00179-0)
- Zegarra-Valdivia, J. A., & Chino Vilca, B. N. (2019). Social cognition and executive function in borderline personality disorder: evidence of altered cognitive processes. *Salud mental*, 42(1), 33–42. [https://doi.org/10.1016/S0378-2166\(02\)00179-0](https://doi.org/10.1016/S0378-2166(02)00179-0)
- Zelazo, P. D., Carter, A., Reznick, J. S., & Frye, D. (1997). Early development of executive function: A problem-solving framework. *Review of general psychology*, 1(2), 198–226. <https://doi.org/10.1037/1089-2680.1.2.198>
- Zhao Q., Guo Q., Hong Z. (2013). Clustering and switching during a semantic verbal fluency contribute to differential diagnosis of cognitive impairment. *Neurosci. Bull.* 29, 75–82. <https://doi.org/10.1007/s12264-013-1301-7>
- Zhao, L., Wang, J. J., & Apperly, I. A. (2018). The cognitive demands of remembering a speaker's perspective and managing common ground size modulate 8- and 10-yearolds' perspective-taking abilities. *Journal of Experimental Child Psychology*, 174, 130–149. <https://doi.org/10.1016/j.jecp.2018.05.013>
- Zhu, N., & Wang, Z. (2020). The paradox of sarcasm: Mentalising and sarcasm use in adults. *Personality and Individual Differences*, 163, 110035. <https://doi.org/10.1016/j.paid.2020.110035>
- Zinck, A., Frith, U., Schönknecht, P., & White, S. (2021). Knowing me, knowing you: Spontaneous use of mentalistic language for self and other in autism. *Autism*, 25(1), 164–175. <https://doi.org/10.1177/1362361320951017>

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological bulletin*, 123(2), 162. <https://doi.org/10.1037//0033-2909.123.2.162>

Zwicker, J., & Müller, H. J. (2013). On the relation between spontaneous perspective taking and other visuospatial processes. *Memory & cognition*, 41(4), 558-570. <https://doi.org/10.3758/s13421-012-0283-4>

