

Social Context Effects on the N400: Evidence for Implicit Theory of Mind?

Alexandra Louise Westley

University College London, Division of Psychology and Language Sciences

A thesis submitted for the degree of Doctor of Philosophy

February 2021

Declaration

I, Alexandra Westley, confirm that the work presented in this thesis is my own.

Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

A Westley

Abstract

A prevailing theory within research into Theory of Mind – the ability to attribute mental states to others – is the existence of a two-system model: an automatic fast-paced “implicit” system, followed by a reason-based “explicit” system. Behavioural evidence supports a double dissociation: children fail social tasks with higher cognitive load but pass implicit measures, and Autistic individuals pass reason-based social tasks with practice but fail implicit measures. This thesis probes the neural basis for implicit Theory of Mind, asking: to what extent do individuals automatically process the comprehension of task partners during a Joint Comprehension task? EEG evidence here indicates participants display an N400 – a neural marker indicating lack of comprehension – when a partner deprived of context cannot understand a sentence displayed, even when the participant has the full context for comprehension. This “Social N400” appears to indicate that participants model partner comprehension, their mental state, in real time. The effect is shown in adults and adolescents; does not appear to be explained by sub-mentalising effects; and notably is absent when the task lacks a prompt to consider the confederate’s comprehension. The results suggest implicit mentalising is not automatic, but a cognitive tool employed when online modelling would aid task demands. The Joint Comprehension task outlined provides a tool to further examine the neural basis of implicit social cognition, particularly within Autism Spectrum Disorder where an impairment in implicit mentalising is suggested.

Impact Statement

Current research into Theory of Mind hypothesises a two-system approach for understanding others' points of view. These are an "implicit" automatic system and an "explicit" reason-based system. It is further hypothesised that Autistic individuals have an impaired implicit Theory of Mind but may make use of explicit Theory of Mind. The nature of a cognitive mechanism underlying any potential implicit system for Theory of Mind is not clear, and evidence for impairment of this system in Autistic individuals is largely behavioural.

This thesis examines early perspective-taking (sub 500 milliseconds) using EEG to establish the potential cognitive basis for implicit Theory of Mind. The results show that participants appear to represent the comprehension of others, in real time, within their own mechanisms for comprehension, termed here a "Social N400". This may also provide an explanatory mechanism for behavioural evidence of developed implicit Theory of Mind in adolescents, as adolescent participants demonstrate comparable online modelling of others to adults. The research further establishes that the Social N400 does not appear to be an artefact of the task and instead may be representative of a fast-paced social perspective-taking mechanism. However, contrary to two-system Theory of Mind, the research finds no support for the hypothesis that such online representation can occur without prompting (automatically), instead suggesting that this is a tool individuals can use when it aids with task demands.

The Joint Comprehension method outlined in this body of research therefore gives future researchers a tool by which to examine early perspective-taking at the neurological level and may lead to a deeper understanding of Theory of Mind in neurotypical individuals. The evidence supports a single system of Theory of Mind, which can employ a number of cognitive tools dynamically based on task demands. This neural evidence may help resolve several inconsistencies shown in behavioural research on Theory of Mind to date.

The research also shows that this task is feasible to conduct with Autistic participants, thereby paving the way for further exploration of the cognitive tools used for Theory of Mind in those on the Autism spectrum. Reliable evidence for differences in early perspective-taking mechanisms may help with diagnostic criteria, and early intervention and support. Indeed, the cognitive marker used here (the N400) is present in infants of 6 months, and researchers modifying the Joint Comprehension paradigm outlined in this thesis have shown similar cognitive results in infants of 14 months (Forgács et al., 2019, 2020). If neural evidence can be reliably combined with behavioural evidence, supportive intervention may be implemented earlier than currently possible.

This body of research was funded by the Economic and Social Research Council and the Medical Research Council.

Data collection for adolescent participants submitted under Chapter 7 was conducted at the University of York in partnership with Dr Shirley-Ann

Rueschemeyer. Analysis of the full data set was conducted at University College London in collaboration with Dr Shirley-Ann Rueschemeyer and Zdenko Kohút at the University of York, and published in *Journal of Experimental Child Psychology* (Westley, A., Kohút, Z., & Rueschemeyer, S.-A. (2017). 'I know something you don't know': Discourse and social context effects on the N400 in adolescents. *Journal of Experimental Child Psychology*, 164, 45–54).

Ethical approval for this body of work was granted by UCL Research Ethics Committee, project ID: 7069/001: Joint action in perception, language, and communication.

To my family, this thesis is testament to your love and support.

Grandad, I finally did it. I love you always.

Table of Contents

Chapter 1	Literature Review.....	15
	Social context in language comprehension.....	16
	Conversation as joint action	24
	Theory of Mind – building a model	28
	The Social N400	35
1.1.1	Can the Social N400 be replicated with a within-subjects design?	37
1.1.2	Can the Social N400 be explained outside of Theory of Mind?	37
1.1.3	Is the Social N400 shown with no task cues	38
1.1.4	Is the Social N400 found in adolescence?	38
1.1.5	Can this work be important in the study of Autism Spectrum Disorder?	
	39	
Chapter 2	The Joint Comprehension Task	41
	Summary	41
	Method.....	43
2.1.1	Participants.....	43
2.1.2	Stimuli	43
2.1.3	Adjustments for adult, neurotypical participants.....	44
2.1.4	Procedure.....	45
2.1.5	EEG Analysis Procedure	47
Chapter 3	The Social N400 in Adults.....	49
	Introduction.....	49
	Methods.....	54

3.1.1	Participants.....	54
3.1.2	EEG Analysis Procedure	54
3.1.2.1	Processing of EEG data	54
3.1.2.2	Canonical N400-Effect.....	54
3.1.2.3	Difference in the N400-Effect across Sentence Conditions	55
3.1.2.4	Latency of N400-Effect	56
	Results.....	56
3.1.3	Behavioural Analysis	56
3.1.4	EEG Analysis	57
3.1.4.1	Canonical N400-Effect.....	57
3.1.4.2	Differences in the N400-Effect across Sentence Conditions.....	57
3.1.4.3	Latency of the N400-Effect	58
	Discussion.....	59
	Chapter 4 Break Effects.....	66
	Introduction.....	66
	Methods.....	70
4.1.1	Participants.....	71
4.1.2	Procedure.....	71
4.1.3	EEG Analysis Procedure	72
4.1.3.1	Processing of EEG data	72
4.1.3.2	Difference in the N400-Effect across Sentence Conditions	72
4.1.3.3	Bayes Factor.....	73
	Results.....	73
4.1.4	Behavioural Analysis	73
4.1.5	EEG Analysis	74
4.1.5.1	Differences in the N400-Effect across Sentence Conditions.....	74

Discussion.....	76
Chapter 5 Object-Based Non-Social Cues	79
Introduction.....	79
Methods.....	82
5.1.1 Participants.....	82
5.1.2 Procedure.....	82
5.1.3 EEG Analysis Procedure	83
5.1.3.1 Processing of EEG data	83
5.1.3.2 Difference in the N400-Effect across Sentence Conditions	83
5.1.3.3 Bayes Factor.....	84
Results.....	84
5.1.4 Behavioural Analysis	84
5.1.5 EEG Analysis	85
5.1.5.1 Differences in the N400-Effect across Sentence Conditions.....	85
Discussion.....	86
Chapter 6 Automatic Implicit Theory of Mind.....	91
Introduction.....	91
Methods.....	94
6.1.1 Participants.....	95
6.1.2 EEG Analysis Procedure	95
6.1.2.1 Processing of EEG data	95
6.1.2.2 Difference in the N400-Effect across Sentence Conditions	95
6.1.2.3 Bayes Factor.....	96
Results.....	96
6.1.3 Behavioural Analysis	96

6.1.4	EEG Analysis	97
6.1.4.1	Differences in the N400-Effect across Sentence Conditions.....	97
	Discussion.....	98
	Chapter 7 The Social N400 in Adolescents.....	102
	Introduction.....	102
	Methods.....	104
7.1.1	Participants.....	104
7.1.2	Stimuli	105
7.1.3	EEG Analysis Procedure	105
7.1.3.1	Processing of EEG data	105
7.1.3.2	Canonical N400-Effect.....	106
7.1.3.3	Difference in the N400-Effect across Sentence Conditions	107
7.1.3.4	Latency of N400-Effect	108
	Results.....	108
7.1.4	Behavioural Analysis	108
7.1.5	EEG Analysis	109
7.1.5.1	Canonical N400-Effect.....	109
7.1.5.2	Differences in the N400-Effect across Sentence Conditions.....	109
7.1.5.3	Latency of the N400-Effect	110
	Discussion.....	111
	Chapter 8 Autism Spectrum Disorders: Single Case Study.....	115
	Introduction.....	115
	Methods.....	117
8.1.1	Participant	117
8.1.2	Stimuli	118

8.1.3	EEG Analysis Procedure	118
8.1.3.1	Processing of EEG data	118
8.1.3.2	Difference in the N400-Effect across Sentence Conditions	118
Results	119
8.1.4	Behavioural Analysis	119
8.1.5	EEG Analysis	119
8.1.5.1	Differences in the N400-Effect across Sentence Conditions.....	119
8.1.6	Task Feasibility.....	121
Discussion	122
Chapter 9 Discussion	125
Limitations of the current task	126
9.1.1	Sample size.....	126
9.1.2	Stimuli	128
9.1.3	“Peeking”	129
9.1.4	Reliance on language.....	130
Developing a model of Theory of Mind	131
9.1.5	Threshold for implicit Theory of Mind.....	134
9.1.6	Implicit vs Explicit Theory of Mind.....	137
9.1.7	Implicit Theory of Mind over the lifespan.....	139
9.1.8	Theory of Mind in Autism Spectrum Disorder.....	141
Conclusion	145

Table of Figures

<i>Figure 1. Example of stimuli used in the Director Task.</i>	18
<i>Figure 2. Example stimuli for the Joint Comprehension Task.</i>	46
<i>Figure 3. Event-related potentials and scalp topography for the Social N400 in adults.</i>	59
<i>Figure 4. Example procedure for the break effects control.</i>	72
<i>Figure 5. Time course of Event-Related Potentials for the break effects control.</i>	75
<i>Figure 6. Time courses of Event-Related Potentials for the non-social control.</i>	86
<i>Figure 7. Time courses for Event-Related Potentials for the Automatic Implicit Social N400.</i>	98
<i>Figure 8. Event-related potentials and scalp topography for the Social N400 in adolescents.</i>	111
<i>Figure 9. Autism Spectrum Disorder case study: Event-Related Potentials and scalp topography for the Social N400.</i>	120

Acknowledgements

The completion of this thesis would have been impossible without the unending advice and encouragement from my supervisor Dr John Swettenham. It has been an utter pleasure working with you. Thank you for always taking the time to listen to my ideas and anxieties, and for pushing me beyond what I thought I was capable of.

Thank you to Dr Jyrki Tuomainen, for your invaluable advice in all things EEG, for reminding me to enjoy the process, and for making me question things in ways I wouldn't have considered.

Thank you to the many members Department of Language and Cognition at UCL who have provided advice and support along the way, to my amazingly hard-working project students, and to my participants. Without you this would not have been possible.

To my Dad, you have been my rock, my sounding board, my sense check. Thank you for helping me be proud of myself.

To my Mom, thank you for believing in me enough for the both of us, and for making sure I always know just how loved I am.

To James, I'm so thankful to have you as my brother. Thank you for prodding me along and putting up with me. We did it, it's done!

Chapter 1 Literature Review

The study of social cognition, psychological processes that support thinking about others' emotions, intentions, and actions (De Jaegher et al., 2010), is multifaceted, with social factors impacting a multitude of forms of human cognition. This has sometimes led to a fractured study, with theories of social cognition developing parallel to one another, but interrogating similar questions. In particular, the current thesis focuses on posited models of social perspective-taking, and the parallels between models of language comprehension.

The Joint Comprehension task that forms the basis of the current thesis has its genesis in the impact of social factors on language comprehension. However, it draws on significant parallels with joint action – where two or more social actors act together to impact their environment, (Sebanz et al., 2006) – and Theory of Mind – the understanding that others have mental states (beliefs, intentions, desires) that are different to our own – (Butterfill & Apperly, 2013). Rather than considering each in isolation, a cross-discipline approach might allow a much deeper understanding of how and when social factors impact individual cognition, and the differences that might be shown across populations.

Social context in language comprehension

Language is a fundamental tool in the human social arsenal; a vessel containing a wealth of social information; and a key component of social interaction. Of the many ways in which language can be thought to facilitate human communication, conversation is one of the most vital social interaction tools (Garrod & Pickering, 2004). Two or more individuals convey messages between one another through use of verbal output, or increasingly with ever-expanding communication devices, written messages. Inherent in this process, individuals must parse information conveyed by their conversation partner, or partners (Brennan et al., 2010; Garrod & Pickering, 2004); they must couch their understanding using both explicitly stated information (Hanna et al., 2003), and information provided implicitly (Van Berkum et al., 2007a); and they must take into account the mental state of their partner (Baron-Cohen et al., 1985). Given the cognitive demands involved in both language comprehension, and in mutual language use, a large focus in research has been on how and when individuals use social information for language comprehension.

A primary window into the comprehension process is to examine how individuals resolve ambiguity in language. Consider the situation in which a colleague asks “can I borrow that book” where you have more than one book on your desk. There are several factors that might help us resolve the ambiguity: perhaps our colleague has a particular area of specialism which

relates to only one book; we might have been discussing a topic relevant to one book within the conversation already; or perhaps our colleague only has a clear view of one of the books on our cluttered desk. A number of these factors relate to common ground, our shared experience, either physically or linguistically. Common ground is therefore a significant piece of social information shared with a conversation partner which may impact language comprehension.

One popular set of tasks to study the impact of social factors on language comprehension is the use of the Director Tasks (Krauss & Glucksberg, 1977). These tasks are specifically used to assess the impact of the physical common ground between individuals on parsing language in the face of ambiguity. In these tasks, a set of shelves with objects placed on them separates two individuals. One individual (the participant) can see all objects on the shelves. The other individual (the Director – a person standing on the other side of the shelves) will have some objects clearly blocked from view. Objects blocked from the Director's view have competitor objects in full view of both participants (for example, two different sized cubes, the smallest is hidden). Throughout the task, the participant is asked to move certain objects by the Director ("move the small cube"). If participants consider the perspective of the Director, they should immediately discount objects which cannot be seen by the Director. However, if participants prioritise their own semantic understanding, they may look to the hidden smallest cube.

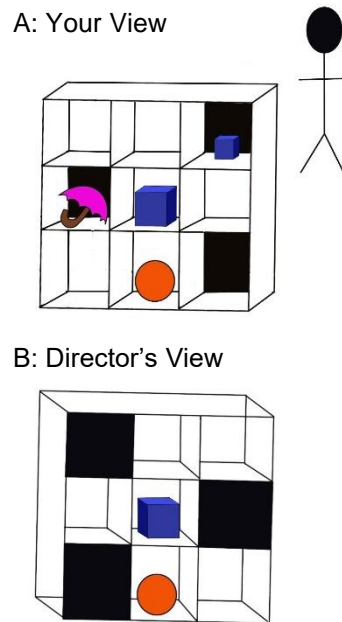


Figure 1. Example of stimuli used in the Director Task. (A: your view, with the Director shown behind the shelves, and B: the view the Director has).

A wealth of evidence shows that in these conditions, participants often look at and reach for objects which are clearly hidden from the Director's view (i.e., not in common ground), and therefore cannot be the object to which the Director is referring (Keysar et al., 2000). This is even shown when the participant themselves is the person who hides the objects in private, and therefore unequivocally knows the Director does not know which items are hidden on the shelves (Keysar et al., 2003).

This evidence has given rise to the theory that individuals first process language from an egocentric point of view, prioritising their own semantic understanding (i.e., the *small* cube must refer to the *smallest* cube despite it being hidden). This initial egocentric interpretation of language is then

reassessed using the context of the shared common ground (Keysar et al., 2000). The theory therefore has two central assumptions: 1) that egocentric language comprehension is a fast and passive stage of comprehension, 2) that use of social context such as common ground is a temporally delayed process.

However, the behavioural evidence from Director Tasks is not consistent, and in many cases stands at odds with this proposed two-stage model of language comprehension. Participants have been shown to be more likely to look at objects in common ground despite reaching for objects in privileged ground (Hanna et al., 2003), indicating that whilst participants are not able to entirely discount objects in privileged ground, common ground does have an immediate impact on sentence comprehension. Further, when the modifier is removed from the instruction (in the case of our prior example, the Director says “move the cube”), child participants look to objects within the common ground early within comprehension and discount the competitor which is hidden (Nadig & Sedivy, 2002), whereas traditional Director Tasks show children tend to show higher failure in reaching patterns when compared to adults (Dumontheil et al., 2010).

Evidence within linguistic common ground is similarly conflicted. One example of linguistic common ground is the precedent set within a conversation with a specific individual. In a task using depictions of ambiguous stimuli, participants were required to locate stimuli referred to by an experimenter. Due to the ambiguous nature of the stimuli, participants and the experimenter built

conceptual pacts, referring to stimuli by specific descriptions through the interaction (for example “the shiny cylinder”). When a new experimenter was introduced and used these descriptions, participants did not immediately look at the associated stimuli the description would have identified if used by Experimenter 1, instead considering multiple referents (Brown-Schmidt, 2009). In this case, the term “shiny cylinder” to refer to a specific ambiguous object was only common ground between the participant and Experimenter 1, suggesting this common ground can have early constraints on language interpretation. This is not, however, the case when the terms within the common ground are given rather than generated through interaction (Kronmüller & Barr, 2007).

Beyond common ground, a variety of social factors are conveyed within a conversation that might be used as context for sentence comprehension. Individuals make use of heuristics across a number of cognitive decision-making functions (Tversky & Kahneman, 1974), and language comprehension is no different. Speech itself, and assumptions made about the speaker, can invoke the use of stereotypes and heuristics (Devine, 1989). Stereotypes contain a wealth of social information, condensed into process-light cognitive shortcuts, and these in turn can have a fundamental impact on the way language is interpreted when in direct contravention to a preconceived stereotype. A sentence such as “I have a large tattoo on my back” spoken by someone with a typical upper-class accent, or “if only I looked like Britney

Spears” spoken by a male, elicit neural markers indicating a lack of comprehension, despite the sentences themselves being semantically plausible (Van Berkum et al., 2007b). The impact of the social context of the speaker can be seen as early as 200-300ms after the beginning of a spoken word. It would seem illogical for complex social information such as stereotypes to impact the earliest stage of language parsing when something as seemingly simply as shared visual space may not.

To reconcile these apparent inconsistencies in use of social context within language comprehension, looking to a model less reliant on temporal stages may provide some answers. For example, a constraint-based competition model (MacWhinney, 1992; Spivey & Tanenhaus, 1998) would propose that multiple cues are available for sentence comprehension and use of each cue is dependent on the cue weight or the resource burden associated. Such a model would not limit the availability of early cues to a static egocentric language process, instead arguing cues distributed across cognitive domains could be available to cooperate and compete based on the task demands (Colvin & Warren, 2020; Skipper, 2015).

Using this lens when interpreting the conflicting evidence from Director Tasks, it could be argued that cue weights could take precedence, or cues could be discarded, based on their computational complexity and resource burden. A key difference between the studies by Keysar et al (2000, 2003) and that by Nadig and Sedivy (2002) is a reduction in the competition between hidden and

visible items. The lowering of this threshold appears to allow for common ground to be used earlier in sentence processing where egocentric reference resolution might otherwise have been favoured.

Tasks which place high burdens on domain-general cognitive constraints may limit the use of cues such as social perspective-taking in sentence comprehension. Executive Functions are higher-order cognitive functions that aid with the ability to think before acting, meet challenges, and control responses (Diamond, 2013). Working Memory (Baddeley, 2003), a capacity-limited store of information for reasoning, learning and comprehension, and Inhibitory Control (Diamond, 2013), the ability to stop an automatic response, are two key components of executive functions relevant to the Director Task which may modulate cue usage. Participants are required to comprehend sentences, keep in mind the common ground, translate comprehension into movement patterns, and suppress mistakes. Indeed, it has been shown that measures of Working Memory and Executive Control are directly correlated with the ability of participants to use perspective-taking to resolve referential statements (Lin et al., 2010; Wardlow, 2013) and that errors in the Director Task decrease from childhood to adolescence, and from adolescence to adulthood (Dumontheil et al., 2010) alongside executive function development. This therefore suggests that errant action or reaching patterns may be due to limitations in participants' ability to use large amounts of contextual cues, or to inhibit the egocentric knowledge of the location of all objects, rather than

perspective information fundamentally not impacting language comprehension.

This is supported by evidence that use of social context as a cue for sentence comprehension is modulated by motivation and reward within the task. In a study where participants had no reward for correct answers in a Director Task, individuals did not appear to incorporate the Director's visual perspective in their anticipatory looking (Cane et al., 2017). When participants were rewarded for speed and accuracy, the Director's perspective was considered at the early anticipatory looking stage, but only when the participants were under a low cognitive load. This again suggests that use of social context in early language comprehension may not be an automatic process, but rather something that can be employed when participants are appropriately motivated; motivation may lower the threshold for the cue being used.

Motivation within tasks may be varied and fall outside of direct reward. In the case of the Kronmüller & Barr (2007) study, participants only used linguistic common ground for early sentence comprehension where the referential pact had been generated through social interaction rather than given. Is the social interaction itself enough motivation to lower the threshold or increase the weight for the use of common ground as a cue? Interpreting conversation as joint action could help to explain the inconsistent use of linguistic common ground as an early sentence comprehension cue, by interpreting collaborative social interaction as a form of motivation.

Less is known about conversation involving more than two participants, and the way this may modulate cue usage. In particular – can other listeners influence the way in which language is interpreted by an individual? What is the impact of another listener having less context to a sentence? Situations in which two individuals are talking, and a third joins at a later point in the conversation are common in everyday social situations from the home to the workplace, but little is understood about how conversation interlocutors both understand and accommodate for any limitations in context a new conversation partner may hold.

Conversation as joint action

Research into the impact of social context on language comprehension can be further explored by considering language, and particularly conversation, as joint action – where individuals coordinate actions to perform a task together (Vesper et al., 2017). In conversation, participants must engage cooperatively with their conversation partners, willingly exchanging information and adjusting based on feedback from others (Barr & Keysar, 2006). Speakers monitor a listener for understanding, and modify their utterances based on feedback; when feedback from listeners is denied, speakers make an increased number of errors in monitoring understanding (Clark & Krych, 2004). In the case of multi-way conversation, i.e., involving more than two individuals, individuals will at points be jointly engaged in the task of listening to, and parsing, the speech of another speaker. In these scenarios, the comprehension of another

listener does not need to have a direct impact on the comprehension of an individual. However, evidence from other joint action tasks indicates that the perspective of others in social tasks, even when it does not directly impact on an individual's own task, can be accommodated for early in processing.

For example, in visual perspective-taking tasks, the visual field of another participant or avatar, has a direct impact on an individual's own task performance, even without prompting. In tasks using avatars, participants view a series of dots on a wall, and are asked to press a button indicating the number of dots present. Whilst this is ongoing, an avatar will also be viewing a series of dots on a wall; the number of these dots may be consistent or inconsistent with the dots in the participant's field of view. Importantly in these tasks, the view of the avatar should have no bearing on the participant's task of saying how many dots are present in the participant's own field of view. Despite this, participants are slower to give correct answers on trials where the avatar can see an incongruent number of dots (Surtees et al., 2016). However, this response can be modulated – the effect can be absent when the avatar is absent for long periods compared to interspersed absent and present. This indicates the effect is social rather than attention direction from the avatar (Ferguson et al., 2017), or perhaps that a social cue carries a higher weight than attention direction.

That another's perspective could impact individual cognition without prompting lends weight to the perspective that this process may be automatic, i.e.,

spontaneous and passive. This is also shown within joint reaction time tasks. In an experiment by Sebanz and colleagues, two participants both completed a Simon task (Craft & Simon, 1970; Simon, 1990), a go no-go task where participants use a left or right button press for a certain condition (e.g. colour); reaction times are faster when the stimulus appears on the side of the screen that corresponds with the correct button press. The two participants completed the task both simultaneously and apart, and it was found that reaction times were faster when the participants were both completing the task together and there was spatial compatibility between the stimulus location and button press for both participants (Sebanz et al., 2003). This effect on reaction times was shown even though the other participant had no direct impact on the individuals' own performance in the task – both were completing the task simultaneously but independently. This suggests that individuals were mentally representing the other participant's task, and this had a direct impact on their own task performance. This effect continues when participants are given the opportunity to ignore the task partner, and to some extent has been shown to impair participants' own task performance (Samson et al., 2010).

In another reaction time task, in which participants were both responding to different dimensions of the same stimulus, it was again shown that the co-action of another participant has a direct impact on cognition, even when not required by the task (Sebanz et al., 2005). Participant reaction times were slower on trials where both participants needed to respond to the stimulus

compared to only one participant responding. This was shown both when participants could see the other participant's action, and when they could not see the action but knew the task of the other participant. The latter result, that the action does not need to be seen, but merely conceptualised, suggests that joint actors are mentally represented, even when task instructions do not call for this as a necessity.

There is some evidence to suggest that the impact of others completing a joint task is rooted directly in imagining the experiences of another, through our own experiences. In an experiment in which individuals guessed how high another participant could jump with weights around their ankles, participants' guesses were improved only when they themselves had experienced moving with the weights (Ramenzoni et al., 2008). This suggests that individuals' perception of the world is rooted within their own capacity for experience. This also works in the reverse, in which observation has a direct impact on subsequent action; when individuals observe another reaching around an object, subsequent reaching patterns are shown to be more curved than without having such an observation, suggesting participants unconsciously represent the impact of the observed obstacle in their action (Griffiths & Tipper, 2009).

Combined, these studies suggest that individuals engaged in task jointly with others will implicitly consider the perspectives of these others. This appears to be the case even when these perspectives do not need to be explicitly considered. It could be the case, therefore, that joint action itself is a motivation

to employ the cue of representing another's perspective. It also appears as if actors and observers share mental tasks across many forms of actions – which lends weight to the potential for actors and observers to share mental tasks within conversation too, as indicated in the modulation of use of common ground by conversational interaction (Kronmüller & Barr, 2007).

Theory of Mind – building a model

Theory of Mind is the understanding that others have mental states – beliefs, intentions, desires – that are different to our own mental states (Baron-Cohen et al., 1985). Theory of mind is believed to be underpinned by mentalising – a process by which an individual represents the mental state of another based on available social cues and context (Green et al., 2015). Some have argued that mentalising can be an implicit operation, assumed to be a fast and automatic process (Frith & Frith, 2012; Green et al., 2015), an idea which aligns with those studies showing that social information can be taken into account in language comprehension in a fast and automatic way.

A popular model within Theory of Mind currently is a temporally bound two-stage model of hypothesising other's mental states:

- 1) an initial fast, automatic process termed “implicit” – where the mental states of others may be represented within the self

- 2) A slower “explicit” process – where Theory of Mind is achieved through reasoning and rule-based application and is heavily reliant on executive functions such as working memory and inhibitory control

The model is underpinned by two key assumptions: that implicit and explicit Theory of Mind are achieved through separate cognitive systems, and that implicit Theory of Mind is engaged spontaneously and without conscious decision (Apperly & Butterfill, 2009; Senju, 2012). Successful use of Theory of Mind in a flexible and dynamic way within everyday social interactions is hypothesised to rely on the coordination of both systems. Support for this model comes from the double-dissociation shown between proposed measures of implicit and explicit Theory of Mind when comparing neurotypical children with children on the Autism Spectrum.

One of the key characteristics of Autism Spectrum Disorder (ASD) is a difficulty in social communication and understanding the mental states of others (Baron-Cohen, 2000). In particular, children on the Autism Spectrum often fail “false belief” tasks such as the Sally-Anne task (Baron-Cohen et al., 1985). In this task – that has been replicated using puppets (Wimmer & Perner, 1983), cartoons (Baron-Cohen et al., 1985), and live actors (Leslie & Frith, 1988) – a child watches as Sally places a marble into a basket, and leaves the room. While Sally is away, Anne moves the marble from the basket to a box. When Sally returns, the child is asked “where will Sally look for her marble?” To correctly pass the task, the child must identify that they hold information about

the location of the marble that Sally would not be privy to, and therefore she will look in the basket, even though the child knows the marble is in the box. Children on the Autism Spectrum will often fail this task, and answer that Sally will look in the box, indicating that they struggle with understanding that Sally holds a mental state different to their own, or with using the contextual cue of Sally's mental state.

In contrast to these findings, it has been shown that there are several situations under which Autistic individuals can learn to pass false belief tasks. Repetition and providing additional reasoning can mean that Autistic individuals learn to pass false belief tasks (Schuwerk et al., 2015) and this is especially true of individuals with no developmental delay in verbal skills (Senju, 2012). These results have been extended to show that children and adolescents on the Autism Spectrum performed equally as well as neurotypical peers in understanding false belief represented through simple stories (Scheeren et al., 2013). In a sample of 194 children and adolescents with a diagnosis of ASD and no developmental delay in verbal skills, participants showed unimpaired understanding of false beliefs, double bluff, faux pas, and sarcasm, as represented in five social stories. However, despite the ability to pass false belief tasks through application of reasoning, children on the Autism Spectrum often fail to apply this reasoning flexibly during everyday social interactions.

Supporting the separation of implicit and explicit Theory of Mind, when false belief tasks are combined with eye tracking, it has been shown that repetition

and learning has differing impacts on verbal outcomes (an explicit measure) and anticipatory looking (an implicit measure). Following an hour of repetition of a false belief task, Autistic individuals showed improved outcomes in correctly passing the task verbally. Despite this, these participants still looked to the incorrect answer, consistent with their own mental state rather than Sally's. Gaze fixation showed no improvement over the hour of repetition for Autistic individuals, in contrast to neurotypical individuals (Schneider et al., 2013). The hypothesis is therefore that the implicit Theory of Mind system is impaired in ASD, but that individuals with strong executive function skills can make use of the explicit system to work out a solution and pass Theory of Mind tasks (Lin et al., 2010). The slower nature, impact of cognitive load, and need to learn social rules could all contribute to the difficulty in applying reasoning about the mental states of others flexibly in social situations.

The other side of the double dissociation comes from evidence from neurotypical development. Children under three years of age often fail a classic explicit test of false belief understanding (Wimmer & Perner, 1983), however they do appear to be able to anticipate the actions of a person who holds a false belief – as demonstrated by anticipatory gaze behaviour recorded on an eye tracker (Southgate, Senju, & Csibra, 2007), a finding that suggests implicit Theory of Mind in young children under three. The current model hypothesises that infants can understand Theory of Mind via implicit means, but the underdevelopment of executive functions means this cannot be

expressed. On some level, implicit and explicit Theory of Mind appear to be independent.

However, there are several key studies which call into question the current model of Theory of Mind and the key assumptions that implicit processes are automatic, and that implicit and explicit processes for Theory of Mind are entirely dissociable. More recent research has shown a failure to replicate evidence from anticipatory looking paradigms (Burnside et al., 2018), which could suggest that implicit Theory of Mind is not always in use, contradicting the automatic assumption. In addition, across other brain processes, it has been shown that we develop shortcuts both cognitively and within brain activation to minimise resource needs for cognition (Skipper & Lametti, 2021; Tversky & Kahneman, 1974); constant activation of implicit processes could be an inefficient use of resources where demand for the process is low.

Implicit processes for Theory of Mind may also be at least partially reliant on the same executive functions that underpin explicit Theory of Mind. The Director Tasks used for models of language comprehension could also be said to involve Theory of Mind for representing the Director's perspective, and the evidence there has already suggested that reducing cognitive load enables more children to pass these types of tests (Nadig & Sedivy, 2002). High cognitive load has also been shown to disrupt looking patterns associated with successful implicit Theory of Mind (Schneider, Lam, et al., 2012). In addition, functional MRI whilst participants watch both false belief, and visual

perspective-taking videos, shows that implicit and explicit Theory of Mind share neural substrates, particularly around the right Temporoparietal Junction (Naughtin et al., 2017; Schurz et al., 2015), and that Theory of Mind brain areas can be shown to activate when viewing social scenarios, even when explicit action is not required.

Looking to models of language comprehension as a comparison, these apparently conflicting results could support a competition model whereby all tools for solving a Theory of Mind problem are available in parallel, and employed based on cognitive constraints, task demands and cue weighting (Brennan et al., 2010). Such a model would not necessitate that implicit Theory of Mind is underpinned by a separate system and is always in use in social situations. Instead, implicit and explicit processes for Theory of Mind could be conceptualised as parallel tools which can be employed where the process is developed, motivation applies appropriate weight to the cue for usage (Cane et al., 2017) and where resources support using the tool as the most efficient process. Theory of Mind itself would therefore be achieved through a multidimensional process, rather than representing a single construct (Warnell & Redcay, 2019).

At the brain level, there is evidence that language comprehension and Theory of Mind share neural substrates, providing support for mutual process models. Language and Theory of Mind networks show synchronised activity at rest and during story comprehension, suggesting a degree of integration between these

brain areas (Paunov et al., 2019). This integration stands up even as Theory of Mind areas are determined using non-verbal tasks. Theory of Mind tasks activate the temporoparietal junction, also a key area for narrative comprehension (Mason & Just, 2009). There is some evidence that a core component of Theory of Mind centres on the amygdala, but that full emergence of Theory of Mind is dependent on a number of co-opted regions underpinning processes such as language, executive functions, and visuo-spatial processing (Siegal & Varley, 2002).

One key gap in current literature is a neural basis for implicit processes in Theory of Mind. It is possible that implicit processes are underpinned by embodiment of others within the self (Gallese, 2007) – i.e. representing or mirroring the perspectives of others within our own brain, the neural equivalent to “putting oneself in their shoes”. The exact mechanism for this embodiment is currently unclear. One suggestion involves mirroring others within neural responses (Gallese & Goldman, 1998) through innate dedicated mirror neurons. Another suggestion is that implicit reasoning may involve bootstrapping from personal perspectives (Goldman, 2006). A neural basis for implicit Theory of Mind would need to explain current behavioural evidence – that the perspectives of others appear to impact the self, that this is shown in children, and that it may not be shown in individuals on the Autism Spectrum.

The Social N400

The N400 is an event related potential response to a lexical or semantic (or possibly pragmatic) violation - for example when a participant reads a sentence such as “The pizza was too hot to run”. The additional processing required to parse an incongruous sentence is shown as a negative voltage spike c.400ms after the onset of the incongruous word. However, when a congruent sentence (without a violation) is presented such as “The pizza was too hot to eat” then the N400 amplitude is smaller. It has been hypothesised that listeners predict upcoming words as a sentence progresses, and when these predictions are violated, additional processing is needed to parse the incongruous word (Kutas & Federmeier, 2011; Mantegna et al., 2019).

Critical to this set of experiments, it has been shown that the N400 can be attenuated if appropriate context is given for the seemingly incongruous utterance, or if the sentence is couched within appropriate context. For example, an animacy violation such as “the peanut was in love” would normally produce an N400, but this effect was shown to be greatly reduced when the sentence was couched within a story about an amorous peanut (Nieuwland & Van Berkum, 2006). The presence of additional supporting context allows listeners to parse more easily what in isolation would be an expectation-violating sentence.

Recently Rueschemeyer et al (2015) have examined whether the N400 response can be shown in a participant when another person (confederate) is presented with a sentence containing a semantic violation. This is achieved by withholding context in a sentence from the confederate but not the participant. When one group of participants heard the supporting context and viewed the potentially incongruous sentence alone, the N400 was attenuated, in line with the findings of Nieuwland and Van Berkum (2006). However, the group of participants who heard supporting context, and then read the potentially incongruous sentences alongside a partner who had this context withheld, showed an N400 response (Rueschemeyer et al., 2015). This response appears to be an electrophysiological marker of semantic integration ‘on behalf of’ or ‘in empathy with’ another person’s lack of comprehension. Hence Rueschemeyer et al (2015) call this N400 a “Social N400” (see Chapter 3 introduction for further detail).

At the outset of this research, the Social N400 had not been replicated by another research group and had only been shown in this between-subjects design. Since then, the Social N400 has been replicated, and has also been shown in a within-subjects but between-conditions design: i.e., participants completed the task alone and also completed the task with a confederate present. Under these conditions, the Social N400 effect occurred when the participant was instructed to adopt the confederate’s perspective, and even without these instructions, which could support the automaticity of implicit

Theory of Mind and that such a tool could be employed even when task motivation is low. However, the Social N400 was not shown when performing a demanding comprehension task, indicating that the process is not entirely dissociable from executive functions, and that it is resource-constrained (Jouravlev et al., 2019).

The current thesis therefore sets out to use Joint Comprehension and the Social N400 to examine a cognitive basis for implicit processes in Theory of Mind.

1.1.1 Can the Social N400 be replicated with a within-subjects design?

In the first study I examine whether the Social N400 effect can be replicated using a within-subjects single-task design. Although between-subjects and within-subjects designs each have their advantages, within-subjects designs have greater statistical power, since random between-subjects variance is eliminated (Field, 2013). A single-task design presents practical advantages in limiting session times and the stimuli pool, and reducing the potential for difference in joint vs alone sessions outside of the presence of a co-reader.

1.1.2 Can the Social N400 be explained outside of Theory of Mind?

It has recently been theorised that many experiments purporting to show Theory of Mind in young infants, and particularly this type of embodiment-based implicit Theory of Mind, are actually confounded by domain-general

“submentalising” (Heyes, 2014). For example, many tasks involve multiple stages, which cause pressure on working memory, or involve a clear attention direction in the form of eye gaze (and can in cases be replicated using a pointing arrow; (Heyes, 2014; Santiesteban et al., 2014, 2015). In this thesis I will examine whether the Social N400 effect is found in response to interruption of the participant (e.g., asking the confederate to open their eyes) (Chapter 4) or in response to the presence of a non-social object such as a camera (Chapter 5).

1.1.3 Is the Social N400 shown with no task cues

A fundamental tenet of the two-system Theory of Mind model is that implicit Theory of Mind is automatic, occurring without prompting when around others. In comparison, a competition-based model of Theory of Mind might posit that where demand for perspective-taking is low, the threshold for using such a tool may not be met. Chapter 6 therefore examines whether the Social N400 effect is shown when there is no prompt to consider the confederate’s perspective.

1.1.4 Is the Social N400 found in adolescence?

Research suggests that the semantic N400 is present from childhood, although the marker may differ in size, latency, and location to its equivalent in adults (Holcomb et al., 1992). For a Social N400 effect to be said to represent implicit Theory of Mind, it must be present prior to adulthood to be used to explain the evidence from behavioural measures of implicit Theory of

Mind. Chapter 7 examines whether the Social N400 effect can be demonstrated in adolescent participants. Given the Joint Comprehension task is heavily reliant on language skills, the presence of the Social N400 is therefore probed in a pre-adult population where previous research indicates adult-equivalent language processing sophistication (Cummings et al., 2008); any differences can therefore be more easily attributed to differences in Theory of Mind. Although children and adolescents show significant development of the brain, and of explicit Theory of Mind and flexible use of social cognition through to adulthood (Blakemore, 2018a; Cummings et al., 2008), implicit Theory of Mind is shown in behavioural measures from infancy (Southgate & Vernetti, 2014).

1.1.5 Can this work be important in the study of Autism Spectrum Disorder?

Given the suggestion that ASD is associated with an impairment in Theory of Mind (Baron-Cohen et al., 1985), and in particular implicit Theory of Mind, it is possible that individuals on the Autism Spectrum may have difficulty representing another person's state of mind in the context of conversation - something which, if the Social N400 can be said to represent implicit Theory of Mind, could be reflected in a reduced effect. The task outlined here is heavily reliant on language comprehension skills, whilst ASD is often associated with language deficits (Grzadzinski et al., 2013). In this thesis I explore the

feasibility of carrying out a study looking at the Social N400 in ASD via a single case experiment.

Chapter 2 The Joint Comprehension Task

The experiments carried out within this thesis use the Joint Comprehension Task developed from the work of Dr Shirley-Ann Rueschemeyer at the Department of Psychology, University of York. The task adapts a previous iteration, initially published as a between-subjects task design (Rueschemeyer et al., 2015), and seeks to bring the task within-subjects to be able to examine multiple linguistic processing effects within the same participant group.

Whilst differences in task design will be made clear in each experimental chapter, a core Joint Comprehension task framework is used throughout the thesis and is therefore outlined here for reference.

Summary

The Joint Comprehension task examines neural recordings gathered using EEG whilst participants and a confederate both attend to written stories on a computer screen.

Participants are seated beside a confederate, who is always a member of the experimental team. At the start of each trial, the confederate is prompted to close their eyes. Whilst the confederate has their eyes closed, the participant silently reads four sentences of a coherent story. After the fourth sentence, the

participant is prompted to ask the confederate to open their eyes, and both jointly read the fifth and final sentence of the story.

The stories used centre on critical nouns, which in all cases in real life would be assumed to be inanimate. The key manipulation is that in half of the stories, the nouns act in a way consistent with their assumed inanimate nature, and in the other half they act as if they are animate. For example, an apple being baked into a pie vs an apple conducting an orchestra. Animacy remains consistent throughout the story in all experimental trials – either the apple is inanimate and being baked, or the apple is animate and enjoying a fulfilling musical career, but it does not switch between states. By having the confederate cover their eyes for the first four sentences of the story, by sentence 5 only the participant has the context that the noun in the story is animate or inanimate.

The task relies on the participant trusting that the confederate is not able to see the first four sentences of each story. To maximise trust, and to minimise the need for the participants to move or divide attention to check whether the confederate's eyes are closed, the confederate gives a visual and movement cue by covering their eyes with their hands and keeping their eyes covered until prompted. It also requires the participant to understand that both parties are reading the final sentence of the story at the same time. The confederate provides a visual and movement cue by uncovering their eyes and leaving their

hands in their lap, but the participant must trust or confirm that the confederate is paying attention to the screen.

Method

2.1.1 Participants

Participants in all experiments self-reported being native speakers of English, with no prior history of neurological impairments, and normal, or corrected-to-normal vision.

2.1.2 Stimuli

Experimental trials consisted of 104 short written stories, i.e., small units of discourse, which were made up of five sentences of 5-9 words ($M = 5.84$, $SD = 1.12$). Each story belonged to one of two conditions (see Figure 2). Plausible (PLAUS) stories were semantically plausible and were made up of 5 sentences which were semantically coherent even when presented alone. Implausible (IMPLAUS) stories presented a coherent narrative across the five sentences, but sentence 1 (S1), sentence 4 (S4) and sentence 5 (S5) were rendered semantically incongruent when read in isolation due to an animacy violation. The first, fourth, and fifth sentences were the critical sentences used for analysis. These followed an agent/verb/object structure. Therefore, incongruent sentences were rendered anomalous only on the presentation of the final word in the sentence (underlined in Figure 2). Each story therefore

contained three target words: the sentence-final words in S1, S4 and S5. This resulted in six experimental conditions: the three target words in the PLAUS condition, and three in the IMPLAUS condition. Target words were identical in all six conditions and were therefore matched for all psycholinguistic variables; these were also chosen with the intention of being simple and easily recognised (mean length = 5.54 letters, SD = 1.42).

To further hone the stimuli, all sentences were rated as plausible or implausible by blind raters (N = 20). The stimuli were rated either in their intended contexts (N = 10), or in isolation (N = 10). PLAUS stimuli were designed to be plausible within a coherent narrative, and each sentence of the narrative should also be plausible in isolation. IMPLAUS stimuli were designed to be plausible within the coherent narrative, but each individual sentence making up the narrative should be implausible in isolation. Stimuli that did not fit these criteria for 2 or more raters were replaced and rechecked.

The full list of stimuli used can be found in the appendix.

2.1.3 Adjustments for adult, neurotypical participants

To reduce detection of the manipulation, 20 stimuli were created as catch trials. Catch 1 trials showed S1-S4 of a PLAUS story, matched with S5 of an IMPLAUS story. Catch 2 trials showed S1-S4 of an IMPLAUS story, matched with S5 of a PLAUS story. All catch trials were removed from final analysis.

2.1.4 Procedure

Participants were seated on a padded immobile chair approximately 70 cm (Chapter 7) or 100 cm (Chapters 3 – 6, 8) from a computer monitor and an EEG cap was fitted. A confederate (one of the experimental team) was seated in an adjacent chair. The participant was led to believe the confederate was another participant completing a separate task after the main experiment. All participants (and parents of adolescent participants in Chapter 7) were introduced to this set-up prior to testing.

Prior to commencing the main experimental task, participants were shown six self-paced examples. The participants were encouraged to ask the experimenter questions and received verbal feedback on their responses. These examples contained two PLAUS, two IMPLAUS, one Catch 1, and one Catch 2 stimuli, in random order. Once the participant understood the task and had completed the examples, the experiment progressed.

Stimuli for the main experiment were presented in 26 blocks of 4 stimuli, with an optional break every 2 blocks. Stimuli were pseudo randomised such that conditions were distributed evenly across the entire experiment, with no condition repeated more than 4 times in a row.

Each trial began with instructions to the confederate to close his/her eyes. Once the confederate's eyes were closed, S1-S4 were presented to the participant on the computer screen (i.e., the confederate could not see these

stimuli with closed eyes). Each sentence was presented a few words at a time across three screens: the first two screens contained 2-4 words each and were visible for 1000ms; the final word of each sentence was presented in isolation for 1000ms. Following the presentation of S1-S4, the participant was prompted to tell the confederate to open his/her eyes. When the participant was satisfied that the confederate's eyes were open, he/she pressed a button, and S5 was then presented on the computer monitor for both participant and confederate to read together.

Following S5, participants answered two questions: Q1 'Do you think the last sentence was plausible for your partner?' and Q2 'Was the last sentence plausible for you?'. Responses were recorded via keyboard press; there was no time limit for this answer. After the final response, a black screen was displayed for 2000ms before the beginning of the next trial. Only correctly answered trials were used in the EEG analysis.

	P reads alone				P&C read together		
	S1			S4	S5	Q1	Q2
PLAUS	The boy rode the <u>bicycle</u> .	The bicycle had a flat tyre.	The boy stopped the bicycle.	The boy got off the <u>bicycle</u> .	The boy pushed the <u>bicycle</u> .	YES	YES
IMPLAUS	The boy argued with the bicycle.	The boy made the bicycle angry.	The bicycle was scared of the boy.	The boy threatened the bicycle.	The boy hurt the bicycle.	NO	YES

Figure 2. Example stimuli for the Joint Comprehension Task. Stories consisting of semantically plausible (PLAUS) and semantically implausible (IMPLAUS) sentences were presented one at a time. The critical target word was the final word in Sentences 1, 4 and 5 (S1, S4, S5) (underlined above). Participant (P) silently views S1-S4 alone. S5 was viewed simultaneously by P and confederate (C). Following each story, P was asked whether S5 was plausible for C (Q1), and whether he/she found S5 plausible (Q2).

2.1.5 EEG Analysis Procedure

For all adult participants, data was collected at the Department of Developmental Science, University College London. Adolescent participant data was collected at the Department of Psychology, University of York, and is outlined separately in Chapter 7 due to differing EEG recording equipment and environment.

Continuous EEG was recorded using BioSemi in a quiet room from 64 shielded active electrodes placed in a 10-10 montage (recording reference = CMS DRL, VEOG and HEOG included, electrode impedances < 25 k Ω , sampling rate = 2024 Hz, resampled to 512 Hz off-line). Data analysis and pre-processing were conducted using EEGLab and ERPLab. Data were re-referenced off-line to the average of the mastoids and filtered using a 0.5 – 30 Hz bandpass filter. Continuous EEG signal was visually inspected for major artefacts, and individual channels with excessive noise were interpolated as an average between two nearest neighbours (Planner interpolation). The signal elicited by

correctly answered trials was segmented (-200 to 1000ms relative to the onset of the target word in each sentence). Trials still containing excessive artefacts were rejected using Moving Window peak-to-peak threshold and Stepwise automatic artefact detection, rejecting gradients $>35\mu\text{v}$, and amplitude spikes $>100\mu\text{v}$. Baseline correction was applied to the time window -200ms to 0ms relative to the onset of the target word. Pre-processed segments were then averaged per condition within participants.

Chapter 3 The Social N400 in Adults

Introduction

Understanding Theory of Mind, and the tools underpinning its successful use, depends on robust methods of investigation. The Joint Comprehension task discussed within Chapter 1 and, most importantly, the Social N400 (Jouravlev et al., 2019; Rueschemeyer et al., 2015) offer the potential for examination of the cognitive mechanisms underpinning the fast-paced Theory of Mind tools or systems termed implicit (Apperly & Butterfill, 2009). The task is designed to treat language comprehension as a joint action (Sebanz et al., 2006) – acknowledging that conversation is an act requiring coordination between individuals (Brennan et al., 2010), and that in order to perfect this act we must always be aware of the comprehension of those we are engaging with (Barr & Keysar, 2006).

The Joint Comprehension task outlined in these previous studies provides individuals with context which enables them to understand implausible sentences and modulates whether participants have a context-deprived confederate present or absent. The initial comprehension task was the same in both Alone and Joint conditions. The participant heard over headphones a context sentence such as “the girl dressed as a canary for Halloween” and then read on a screen a context-dependent sentence such as “the girl had a little beak” (Jouravlev et al., 2019). When the participant read “the girl had a

little beak” with no partner present (Alone), they displayed no N400 which would indicate a lack of comprehension, in line with previous research showing context can be used as a cue for comprehension early in language processing (Nieuwland & Van Berkum, 2006). However, for Joint sessions there was a confederate present who had no headphones, and therefore was not provided with the context sentence. When the participant and confederate read the sentence from the screen, participants showed an N400 in the window after the onset of the word “beak”, indicating a lack of comprehension termed a Social N400. Participants showed high accuracy in their judgements of plausibility for themselves and for the confederate, indicating simultaneous processing of both perspectives (Jouravlev et al., 2019; Rueschemeyer et al., 2015). From these results, we might infer that all participants can use semantic context early in comprehension, however when there is a cue to consider a jointly comprehending partner, participants model the partner’s comprehension early whilst also comprehending a sentence for themselves.

Bringing this task within-subjects and to a single-task design offers the opportunity to show both presence and absence of the N400 within the same group of participants within the same task – i.e., use of both semantic context and social context by the same individuals. To design such a task, it’s necessary to understand what is underlying the presence and absence of the N400 marker, and therefore what counts as context for the brain when comprehending language. The N400 itself has been shown to be a marker for

a mismatch with expectation, and can be shown across many situations, even outside of language (Kutas & Federmeier, 2011). Within language, it is hypothesised that individuals predict upcoming words within sentences (Van Berkum et al., 2005). If we take the example from the Jouravlev study, “the girl had a little...” readers generate expectations for the final word – perhaps “teddy” or “toy”. The N400 therefore could represent the extra cognitive resource required to undo expectation and fit the word “beak” into the sentence. With the additional context of “the girl dressed up as a canary for Halloween” – as participants read the sentence “the girl had a little...” then “beak” becomes more primed for expectation than without that context (Nieuwland & Van Berkum, 2006). This therefore suggests that the key impact of context on the presence or absence of an N400 is whether the context supports the prediction of a word within a sentence.

Bringing the Joint Comprehension task within subjects therefore hinges on the expectation of words within a sentence. We designed five sentence stories where in isolation the prediction of the final noun in each sentence is modulated by expectations about animacy: plausible vs implausible. However, the same noun is used consistently throughout the story. This means that on sentence 1 participants could show an N400 based on violation of expectation in an implausible story – referring to the example in Figure 2 “the boy argued with the...” would not support prediction of the noun “bicycle” (implausible), whereas “the boy rode the...” would support the prediction of “bicycle”

(plausible). However, the bicycle remains consistent throughout the story. By sentence 4, even in the implausible “the boy threatened the...” the participant is primed to anticipate the noun “bicycle” based on the context of all the previous sentences they have read, and therefore we would not expect to see a significant N400 when compared to the plausible “the boy got off the bicycle”, in line with previous language studies (Van Berkum et al., 2005).

Our primary investigation here is what happens when a participant reads sentences 1 – 4, a confederate is not able to read these sentences, but both participant and confederate read sentence 5. In this situation, the participant themselves should still be primed to expect the noun “bicycle” in the previous example. However, the confederate is not primed to expect this noun, and would need to dedicate additional cognitive resources to re-examine their expectation and fit “bicycle” into an implausible sentence compared to a plausible sentence, the same as the participant had to do on sentence 1. If a participant displays a Social N400, this could indicate that they are no longer prioritising the context of the previous four sentences to prime their own expectations, but instead are reading the sentence as if they are their context-deprived reading partner, simulating the partners’ comprehension using their own cognitive mechanisms for understanding.

Creating a within-subjects Joint Comprehension design has two key benefits. Firstly, the task itself holds greater statistical power as variances between participant groups are eliminated (Field, 2013). For example, in the

Rueschemeyer et al (2015) study, was there an unmeasured systematic difference between those in Group 1 comprehending alone, and those in Group 2 with a partner present, that might otherwise explain the Social N400. Bringing the test to a single task also controls for potential systematic differences between Joint and Alone sessions. Beyond this, it also provides a method for interrogating the Social N400 as a neural indicator of implicit Theory of Mind through future task modulations. Representing a reading partner's mental state (their comprehension) in a time frame sub 500 milliseconds could indicate fast-paced implicit Theory of Mind (Apperly & Butterfill, 2009), and robustly interrogating this could allow us to understand this cognitive phenomenon to a greater extent.

The current experiment therefore aimed to test a within-subjects single-task version of the Rueschemeyer et al. (2015) Joint Comprehension task in adult participants. The experiment aims to demonstrate the use of semantic integration for the self, and the simulation of the perspective of a jointly comprehending partner, within the same participants. Based on previous investigations of context effects on comprehension (Nieuwland & Van Berkum, 2006; Van Berkum et al., 2005), we hypothesised that adults would integrate semantic information across short stories, showing an N400 effect for implausible stories at sentence 1, which we expected would attenuate by sentence 4. We hypothesised that adult participants would integrate the

comprehension of the jointly attending confederate, showing a Social N400 effect on sentence 5 of implausible stories.

Methods

The full methods for the Joint Comprehension task are outlined in Chapter 2.

3.1.1 Participants

A total of 20 participants aged between 19 and 61 years were recruited through University College London (12 men; median age = 20 years, 3 months; SD = 11 years, 10 months). Five participants were removed from the final analysis due to too few trial contributions following excessive noise in the EEG signal. Therefore, 15 participants remain in the final analysis (9 men; median age = 20 years, 3 months; SD = 13 years, 11 months).

3.1.2 EEG Analysis Procedure

3.1.2.1 Processing of EEG data

Following processing of EEG data as outlined in Chapter 2, the average number of trials included in the analysis was S1 PLAUS = 30, S1 IMPLAUS = 29, S4 PLAUS = 30, S4 IMPLAUS = 29, S5 PLAUS = 31, S5 IMPLAUS = 30.

3.1.2.2 Canonical N400-Effect

To determine the time window in which reliable differences between PLAUS and IMPLAUS stimuli were seen on S1 (i.e., a canonical N400-Effect), the ERP

signal from these conditions were submitted to a repeated measures, two-tailed permutation test based on the tmax statistic (Blair & Karniski, 1993) using a family-wise alpha level of 0.05 (Bullmore et al., 1999; Groppe et al., 2011) using the Mass Univariate ERP Toolbox. All time points between 200 and 500ms post presentation of the critical word in each condition on S1 at 9 key scalp electrode sites were included in the test (F3, Fz, F4; C3, Cz, C4; P3, Pz, P4) following precedent from previous investigations using this experimental procedure (Rueschemeyer et al., 2015). Repeated measures t-tests were performed for each comparison using the original data and 2500 random within-participant permutations of the data, over twice the number recommend (Manly, 1997). The most extreme t-score in each of the 2501 sets of tests (i.e., the "tmax" of each set of tests) was recorded and used to estimate the tmax distribution of the null hypothesis (i.e., no difference between conditions). Based on this estimate, critical t-scores of +/- 4.51 (df = 14) were derived. In other words, any differences in the original data that exceeded a t-score of +/- 4.51 were deemed reliable, using an alpha level of 0.000486 to correct for multiple comparisons.

3.1.2.3 Difference in the N400-Effect across Sentence Conditions

To compare the N400 across sentence conditions, the mean amplitude of the ERP signal from the nine scalp electrode sites was extracted from each participant for each of the six experimental conditions (PLAUS, IMPLAUS at each of the 3 sentence positions), based on the time window shown to display

a robust Canonical N400 on S1 (350-400ms). To assess whether significant differences between the conditions could be seen across sentence positions, mean amplitude changes were first entered into a 2x3x9 repeated measures ANOVA with the factors Condition (PLAUS, IMPLAUS), Sentence (S1, S4, S5), and Electrode (F3, Fz, F4; C3, Cz, C4; P3, Pz, P4). Following this, amplitude of the N400 at each target time point (S1, S4, S5) was calculated using the difference in amplitude between PLAUS and IMPLAUS, averaged over the midline (Fz, Cz, Pz). Planned directional (p -value halved) one sample t -tests (Holm's Sequential Bonferroni Procedure used to account for multiple comparisons) were run on mean amplitudes to test for a significant N400 in each sentence position.

3.1.2.4 Latency of N400-Effect

In addition, the peak latency of the N400-Effect elicited on S1 and S5 was extracted and entered into a paired samples t -test in order to test for differences in the latency of the effect across sentence positions.

Results

3.1.3 Behavioural Analysis

Responses to the two questions posed after each trial were analysed in a 2 x 2 repeated measures ANOVA with Question (Q1, Q2) and Condition (PLAUS, IMPLAUS) as main factors. Accuracy rates were high (Q1-PLAUS: $M = 98.08\%$, $SD = 2.18\%$, Q1-IMPLAUS: $M = 97.39\%$, $SD = 3.38\%$, Q2-PLAUS:

M = 98.78%, SD = 1.59%, Q2-IMPLAUS: M = 95.64%, SD = 4.99%), indicating that participants remained engaged in the task. There was no main effect of Question ($F(1, 14) = 1.47, p = .25$), Condition ($F(1, 14) = 2.06, p = .17$), or interaction between Question and Condition ($F(1, 14) = .18, p = .68$) indicating that both questions were responded to with equal accuracy, and both conditions had similar levels of correct interpretation.

3.1.4 EEG Analysis

3.1.4.1 Canonical N400-Effect

The tmax permutation analysis revealed a small significant time window from 350-400ms ($p < 0.001$). In this time window, IMPLAUS target words elicited a significantly stronger negative signal than PLAUS target words. The time window and topography of this effect is broadly consistent with those generally seen for the N400-Effect (see Figure 3) although it maintains an earlier time window and somewhat more central distribution than previous literature indicates.

3.1.4.2 Differences in the N400-Effect across Sentence Conditions

Mean amplitude of the ERPs elicited by the target words in all six conditions (S1, S4, S5 for both PLAUS and IMPLAUS trials) were calculated in the time window defined by the S1 N400-Effect (350-400ms) and entered into a 2x3x9 repeated measures ANOVA with the factors Condition (PLAUS, IMPLAUS), Sentence (S1, S4, S5), and Electrode (F3, Fz, F4; C3, Cz, C4; P3, Pz, P4).

Main effects of both Condition, $F(1,14) = 8.71$, $p = .011$, $\eta^2 = .38$, and Sentence, $F(2,28) = 31.29$, $p < .001$, $\eta^2 = .69$, were observed. In addition a significant Condition x Sentence x Electrode interaction was observed, $F(16,224) = 1.72$, $p = .044$, $\eta^2 = .11$, indicating that plausibility affected the signal differently at the different sentence positions. The interaction with electrode is broadly consistent with a single component varied in amplitude across conditions (Luck, 2014). Planned comparisons between the plausibility conditions (IMPLAUS < PLAUS) at each sentence position demonstrated a significant effect on S1, $t(14) = 3.01$, $p = .004$ and S5, $t(14) = 2.31$, $p = .019$, but not on S4, $t(14) < 1$, $p = .242$).

3.1.4.3 Latency of the N400-Effect

There was no difference between the peak latency of the N400-Effect elicited on S1 (Mean = 387ms, SD = 27ms) and S5 (Mean = 369ms, SD = 35ms), $t(14) = 1.49$, $p = .159$.

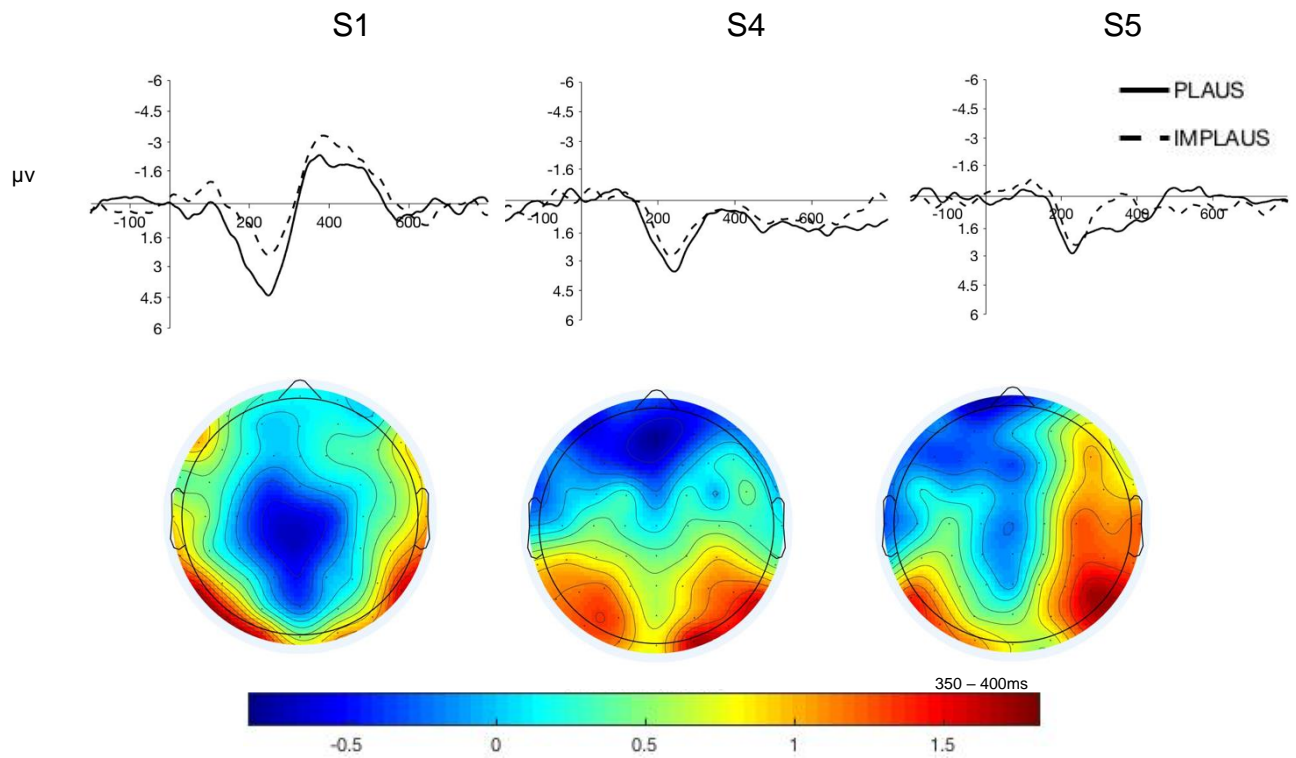


Figure 3. Event-related potentials and scalp topography for the Social N400 in adults. Time courses of ERPs can be seen for target words in each sentence position (S1= Sentence 1, S4= Sentence 4, S5 = Sentence 5). ERPs elicited by target words in semantically plausible sentence stimuli are depicted by the solid line; ERPs elicited by target words in semantically implausible sentence stimuli are depicted by the dashed lines. In the bottom panel scalp distributions showing differences between the conditions across the N400 time window are seen for each sentence position.

Discussion

In the current study, we investigated use of a within-subjects single-task Joint Comprehension task using adult participants. This task assesses the impact of context on semantic integration and the effect of the perspective of a jointly attending partner. Our results suggest that adults make use of context

information in semantic integration, using semantic context to prime expectations. Adult participants showed a clear N400 (indicating incongruence) on sentence 1 (S1), which attenuated with additional context repetition, until no reliable N400 is shown at sentence 4 (S4). Our results also suggest that adult participants simulated the comprehension of a confederate who was context-naïve and jointly attending for sentence 5 (S5), showing a Social N400 effect despite demonstrably understanding sentences for themselves.

The key future intention for this paradigm, is to be able to use Joint Comprehension and the Social N400 to investigate Theory of Mind. In particular, to ascertain whether the Social N400 is indicative of an implicit Theory of Mind - hypothesised to rely on the automatic embodiment of another within one's own cognitive systems, in order to understand other perspectives (Apperly & Butterfill, 2009; Gallese, 2007). The current results appear to indicate adult participants used real-time simulation of the perspective of their context-naïve partner, which could be described as implicit Theory of Mind. Specifically, participants should be primed to anticipate the critical noun in sentence 5, based on the four sentences they have read alone. However, participants appear to model the comprehension of the confederate, who has not been primed to anticipate the critical noun and would therefore need to expend additional resource to re-evaluate expectation and parse the critical noun into the sentence. To claim this, we need to examine whether the Social

N400 truly reflects a) an equivalence to one's own cognitive systems for understanding, and b) a true reflection of social perspective taking.

The results support the interpretation that the Social N400 effect indicates participants representing the comprehension of others within their own cognitive systems for comprehension. To claim this, the Social N400 on S5 must be equivalent to the canonical N400 shown on S1. In adults the N400 is an effect detected towards the posterior of the scalp, and effectively demonstrated using an average of the midline electrodes (Kutas & Federmeier, 2011; Rueschemeyer et al., 2015). We therefore used the midline as the cluster of interest for analysis for both the S1 N400 and the S5 Social N400, such that the N400 would need to be topographically equivalent for each. The original Rueschemeyer study identified the time window for analysis in their original experiment based on eyeballing the waveform plot; to reduce bias in the current experiment we used a two-tailed permutation test based on the tmax statistic to determine the time window used for analysis. This was conducted using the N400 on S1, and this same time window was applied to analysis for the Social N400 on S5 such that the effect would need to be shown within equivalent time windows. Further interrogation confirmed no significant difference in peak latency between the S1 N400 and the S5 Social N400. The results therefore suggest that the N400 effects seen on S1 and S5 are similar in topography and time window, supporting the theory that participants are representing others within their own cognitive systems for understanding.

An interesting result to note from our results for this experiment is that the Social N400 shown on S5 appears to be less negative than the canonical N400 shown on S1. This represents a deviation from the results of the Rueschemeyer et al (2015) study, as in the original experiment both the N400 and Social N400 showed similar forms. However, the current study made a critical change in modifying the paradigm to a within-subjects design. To build context for the participant, and then have the final shared sentence with the confederate, the participants were reading five-sentence stories. The sentence structure for each was such that the same noun was the subject of each sentence in the story (e.g., each sentence of the story was about a bicycle). Thus, by S5, participants were primed to anticipate the critical noun (e.g., bicycle). Previous research shows that when a stimulus is primed, the N400 becomes more positive (Bentin et al., 1985; Zhang et al., 2006). Therefore, the difference in the positivity of the N400 between S1 and S5 is in line with expectation; at S1 the object is not yet anticipated, whereas by S5 it has been primed across four prior sentences.

To provide a useful tool for understanding Theory of Mind, the N400 on S5 must robustly represent the impact of social context, above and beyond explanations termed “submentalising” (Heyes, 2014; Santiesteban et al., 2015). For example, reliance on word repetition for context priming could raise the question as to whether the effect described can be attributed simply to habituation. Participants may show an N400 on S1 which is attenuated by S4

due to the critical word becoming habituated – the neural response suppressed (Huber & O'Reilly, 2003; Jacob & Huber, 2020; Rieth & Huber, 2017). This then raises questions regarding the break between S4 and S5, in which the participant is prompted to tell the confederate to open their eyes. This break, whilst small, may present a window for dishabituation – allowing the neural response to increase. The break also has potential to introduce memory artefacts – participants may show a resurgence of the N400 on S5 simply because the context has moved from short term memory and therefore is no longer being used to couch interpretation.

Another potential explanation outside of social cues is that whether the confederate is looking at the screen or not represents a change in attention cues. The confederate directing their visual attention to S5 may lead to resurgence of the N400 simply due to additional attention direction. For example, in visual perspective-taking tasks, effects of a partner's visual attention supposed to be social in nature can in fact be replicated by replacing the partner with an arrow pointing in the direction of the stimuli (Santesteban et al., 2014). To ascertain the veracity of the apparent Social N400 effect, in the following chapters we will aim to investigate whether either of these artefacts could produce the N400 on S5 without it being related to a social cue.

The current study uses a relatively small, although not uncommon within EEG research, sample size. The sample size was determined by precedent within published research into the N400, given the large size of the effect (Bhavnani

et al., 2021). More recently, however, there has been a greater awareness of the strength of significant effects and increasing the robustness for selection of sample size (Larson & Carbine, 2017). The effects outlined here are significant but using power analysis to determine effect size going forward could increase the validity and comparability across studies, which can be limited when sample sizes are small (Baker et al., 2021). This is discussed in greater detail in Chapter 9 as an overall critique of this body of work.

Variations of the current paradigm could also be used to build a model of Theory of Mind. A key tenant of implicit Theory of Mind within the two-system model is the spontaneous nature with which individuals are hypothesised to model others (Gallese, 2007; Gallese & Goldman, 1998; Schurz et al., 2015), whereas other models might predict that implicit Theory of Mind is used as a tool when it would aid task demands (MacWhinney, 1992; McRae & Matsuki, 2013; Spivey & Tanenhaus, 1998). In our study, participants were prompted to consider the perspective of their jointly attending partner. It is likely that participants soon learn to expect a question asking them to explicitly consider what their partner was thinking, as each trial was succeeded by the question “would the final sentence have been plausible to your partner?” Therefore, we cannot be sure that the Social N400 displayed in our results would be automatically shown in participants without explicit prompting. In Chapter 6, we move to investigate whether the Social N400 is present when participants

are not prompted to consider the perspective of their partner, and we evaluate what this might mean for a model of Theory of Mind.

The study here provides a paradigm to investigate online simulation of the perspectives of others jointly attending to language. Adult participants appear to model the comprehension of a jointly attending partner online, within their own language comprehension systems. This modelling occurs in real-time, with adults displaying a Social N400, despite parsing the sentence correctly for themselves. The results therefore provide key control data, from which we can move to control key factors, such as spontaneity of perspective-taking, memory interference, and attention artefacts, and relate the current results to true implicit Theory of Mind.

Chapter 4 Break Effects

Introduction

In the previous chapter, we used a within-subjects Joint Comprehension task to show that adult participants appear to model the comprehension of co-reading partners online. That is, the marker for semantic incongruence, the N400, attenuates when participants are presented with increasing amounts of semantic context throughout a written narrative. When a context-naïve partner jointly reads the final sentence of this narrative, participants appear to represent the partner's lack of comprehension within their own language systems, in real time, showing an apparent Social N400. Whilst this paradigm shows promising use for exploring representation of others within the self for perspective-taking (mentalising), a growing body of literature outlines that many key paradigms purporting to show implicit mentalising, could in fact be explained using domain-general submentalising processes (Conway et al., 2017; Ferguson et al., 2017; Heyes, 2014; Santiesteban et al., 2014, 2015).

Heyes (2014) argues that what is described as implicit mentalising in fact represents domain-general cognitive mechanisms (such as attention or memory) that lead to behaviour appearing to involve mentalising whilst not actually involving the representation of another's mental state. For example, in an experiment purporting to demonstrate implicit mentalising (Senju et al., 2009) adult participants watched a typical Sally-Anne false belief scenario (as

described in Chapter 1). The critical test is whether the participant (whose eyes are being tracked) anticipates the human agent's actions (i.e., that she will reach toward the empty box she believes the ball is in), by first looking at the empty box and looking there for longer than the other box. The results showed that participants did indeed look towards the empty box, and this was interpreted as evidence of implicit Theory of Mind. However, Heyes (2014) argues that participants may simply have been distracted by the human agent's head movements at the critical moment in the film (agent looking away), leading to the participant failing to remember the most recent movement of the ball and simply focusing their gaze at the box they believed the ball would be in. Thus, Heyes suggests that before we can be confident that an experiment genuinely involves implicit mentalising, we need to control for submentalising artifacts.

The current paradigm contains two potential key artefacts – break effects, and attention orienting. The following two chapters aim to address these key artefact potentials.

One potential reason for a resurgence in the N400 when the partner jointly reads the final sentence, could be due to the divided nature of the procedure. Firstly, the participant reads the first 4 sentences of the narrative alone while the partner has closed eyes. The participant then pauses to tell their partner to open their eyes, and both jointly read the final sentence. This has the potential to introduce two key artefacts due to the nature of human memory: retroactive

interference (Heyes, 2014), and event boundary effects (Radvansky et al., 2011). It also provides a potential window for dishabituation if the attenuation between S1 and S4 is explained by dampening of neural response to the repeated target word (Rieth & Huber, 2017).

The inclusion of a break during which the partner is prompted to open their eyes, represents a salient event during the paradigm. It notes a key departure from the continuation of the four previous sentences within the stories. This event has the potential to cause retroactive interference (Heyes, 2014), i.e., poor memory for the preceding event (in this case, the previous context for the final sentence). Retroactive interference occurs when later information disrupts the memory of previous information: for example, participants given one list of words to memorise, and then a second list of words to memorise, remembered significant fewer of the words on list one compared to participants only given a single list to remember (Melton & Lackum, 1941). Remembering to break and prompt the confederate to open their eyes could therefore cause the context from S1 to S4 to have less of an effect on the interpretation of the final sentence for the participant themselves – meaning the N400 shown on S5 is not social in nature, but represents the participants own understanding.

Similarly, the break could also represent a boundary within the task. Boundary events have been shown to cause event segmentation, during which a new event model is created and stored in memory (in this case S5), and the previous event model gradually becomes less available (in this case S1 – S4)

(Kurby & Zacks, 2008; Swallow et al., 2009). For example, a common occurrence is for individuals to forget what they have entered a room to look for once they cross the threshold of the door – with the door encoded within the memory as a boundary event, reducing availability of the memory of what needs to be searched for (Radvansky et al., 2011). Like retroactive interference, this would reduce the availability in memory of the preceding context information and could mean that the N400 seen on S5 represents the participants own understanding, rather than social modelling.

Outside of memory, the break combined with the repetitive nature of the stimuli used, could mean an N400 on S5 is due to dishabituation. The same critical noun is used in all five sentences of each stimulus story. In the usual Joint Comprehension task, the participant sees this noun repeated four times prior to the break. There is therefore potential that attenuation of the N400 from S1 to S4 is due to habituation of response to the critical noun, as repeated letter strings show greater habituation (Rieth & Huber, 2017). The break prior to S5 could provide a window for dishabituation, such that the N400 on S5 is due to greater neural response to the critical noun for the participant themselves, again a non-social effect.

To address this, the current experiment aims to replicate both the building and repetition of context over time, and the break related to giving an instruction to a partner, whilst removing the joint comprehension aspect of S5. This should therefore mean that any N400 seen on S5 is due to the participant lacking

semantic comprehension, and therefore we can infer that the break has caused an artefact, interfering with the use of the previous context from S1 – S4.

Specifically, in the current experiment, the participant and the partner will read S1 – S4 jointly. The participant will then be prompted to ask the partner to cover their eyes, and the participant will then read S5 alone. To minimise difference to Chapter 3, and maintain the ability to compare the results, the participant will be asked two questions after each trial, regarding both their own and the partners comprehension – Q1 *‘Do you think the last sentence would have been plausible for your partner?’* and Q2 *‘Was the last sentence plausible for you?’*. This change in the procedure means that there is no joint comprehension for S5, and that even if the participant considers the partners perspective when reading S5, the partner has full context. The partner serves only to create potential memory interference when the participant breaks to ask the partner to close their eyes. The starting position for this investigation is that the Social N400 is exactly that – social in nature. Therefore, we would anticipate no N400 on S5 if there are no significant artefact effects of breaking between S4 and S5.

Methods

The full Joint Comprehension task is outlined in Chapter 2 and is used here with the key modification outlined in 4.1.2.

4.1.1 Participants

A total of 19 participants aged between 18 and 52 years were recruited through University College London (6 men; median age = 25 years, 9 months; SD = 9 years, 1 month). Five participants were removed from the final analysis due to too few trial contributions following excessive noise in the EEG signal. Therefore, 14 participants remain in the final analysis (3 men; median age = 23 years, 10 months; SD = 10 years, 6 months).

4.1.2 Procedure

To maximise comparison between the current experiment and that outlined in Chapter 3, the procedure used was maintained (as in Chapter 2) except for the key manipulation highlighted below.

Each trial began with instructions to the confederate to *open* his/her eyes. S1-S4 were presented to both the participant and the confederate on the computer screen. Following the presentation of S1-S4, the participant was prompted to tell the confederate to *close* his/her eyes. When the participant was satisfied that the confederate's eyes were closed, he/she pressed a button, and S5 was then presented on the computer monitor for the participant to read alone (i.e., the confederate did not see S5).

	P & C read together				P reads alone		
	S1			S4	S5	Q1	Q2
PLAUS	The boy rode the <u>bicycle</u> .	The bicycle had a flat tyre.	The boy stopped the bicycle.	The boy got off the <u>bicycle</u> .	The boy pushed the <u>bicycle</u> .	YES	YES
IMPLAUS	The boy argued with the bicycle.	The boy made the bicycle angry.	The bicycle was scared of the boy.	The boy threatened the bicycle.	The boy hurt the bicycle.	NO	YES

Figure 4. Example procedure for the break effects control. Stories consisting of semantically plausible (PLAUS) and semantically implausible (IMPLAUS) sentences were presented one at a time. The critical target word was the final word in Sentences 1, 4 and 5 (S1, S4, S5) (underlined above). S1-S4 were viewed by both the participant (P) and confederate (C). S5 was viewed silently by P alone. Following each story, P was asked whether S5 would have been plausible for C (Q1), and whether he/she found S5 plausible (Q2).

4.1.3 EEG Analysis Procedure

4.1.3.1 Processing of EEG data

The average number of trials included in the analysis was S1 PLAUS = 33, S1 IMPLAUS = 34, S4 PLAUS = 35, S4 IMPLAUS = 32, S5 PLAUS = 34, S5 IMPLAUS = 34.

4.1.3.2 Difference in the N400-Effect across Sentence Conditions

As no change was expected to the N400 on S1 between this experiment and that outlined in Chapter 3, the time window was not recalculated. To directly compare the current results to the Social N400 effect measured in Chapter 3,

the mean amplitude of the ERP signal from the midline scalp electrode sites was extracted from each participant for each of the six experimental conditions (PLAUS, IMPLAUS at each of the 3 sentence positions), based on the time window shown to display a robust Canonical N400 on S1 in Chapter 3 (350-400ms). Amplitude of the N400 at each target time point (S1, S4, S5) was calculated using the difference in amplitude between PLAUS and IMPLAUS, averaged over the midline (Fz, Cz, Pz; in line with Chapter 3). Planned directional (p -value halved) one sample t -tests (Holm's Sequential Bonferroni Procedure used to account for multiple comparisons) were run on mean amplitudes to test for a significant N400 in each sentence position, as in Chapter 3.

4.1.3.3 Bayes Factor

The critical outcome of this study predicts a null result for S5 if the break in the task does not present a task artefact. To further test this, amplitude of the N400 on S5 was entered into a Bayesian One-Sample t -test to contrast the likelihood of the data fitting under the null hypothesis with the likelihood of fitting under the alternative hypothesis.

Results

4.1.4 Behavioural Analysis

Responses to the two questions posed after each trial were analysed in a 2 x 2 repeated measures ANOVA with Question (Q1, Q2) and Condition (PLAUS,

IMPLAUS) as main factors. Accuracy rates were high (Q1-PLAUS: $M = 97.38\%$, $SD = 4.00\%$, Q1-IMPLAUS: $M = 97.91\%$, $SD = 3.81\%$, Q2-PLAUS: $M = 93.36\%$, $SD = 6.68\%$, Q2-IMPLAUS: $M = 94.94\%$, $SD = 6.31\%$), indicating that participants remained engaged in the task. There was no main effect of Question ($F(1, 13) = 1.69$, $p = .216$). However there was a main effect of Condition ($F(1, 13) = 13.32$, $p = .003$). There was no significant interaction between Question and Condition ($F(1, 13) = 1.44$, $p = .252$) indicating that both questions were responded to with equal accuracy, however the IMPLAUS condition had lower levels of correct interpretation.

4.1.5 EEG Analysis

4.1.5.1 Differences in the N400-Effect across Sentence Conditions

In order to make direct comparison to the analysis in Chapter 3, mean amplitude of the ERPs elicited by the target words in all six conditions (S1, S4, S5 for both PLAUS and IMPLAUS trials) were calculated in the time window 350-400ms and entered into a $2 \times 3 \times 9$ repeated measures ANOVA with the factors Condition (PLAUS, IMPLAUS), Sentence (S1, S4, S5), and Electrode (F3, Fz, F4; C3, Cz, C4; P3, Pz, P4). Main effects of both Condition, $F(1,13) = 5.04$, $p = .043$, $\eta^2 = .23$, and Sentence, $F(2,26) = 25.72$, $p < .001$, $\eta^2 = .66$, were observed. In addition there was a significant main effect of Electrode with Greenhouse-Geisser correction, $F(2.29,29.78) = 5.50$, $p = .007$, $\eta^2 = .30$. The main effect of electrode is broadly consistent with a single component varied in amplitude across conditions (Luck, 2014). The interaction between

Condition x Sentence x Electrode was not significant, $F(16,208) = 0.72$, $p = .769$, $\eta p^2 = .05$. Planned comparisons between the plausibility conditions (IMPLAUS < PLAUS) at each sentence position demonstrated a significant effect on S1, $t(13) = 1.89$, $p = .040$, but not on S5, $t(13) = 1.64$, $p = .063$ (Holm's Sequential Bonferroni $\alpha = .025$) or S4, $t(13) = .10$, $p = .463$. An estimated Bayes factor (null/alternative) for N400 amplitude on S5 suggested the data were 1.56:1 in favour of the null hypothesis, or 1.56 times more likely to occur under a model without including break effects compared to one including them.

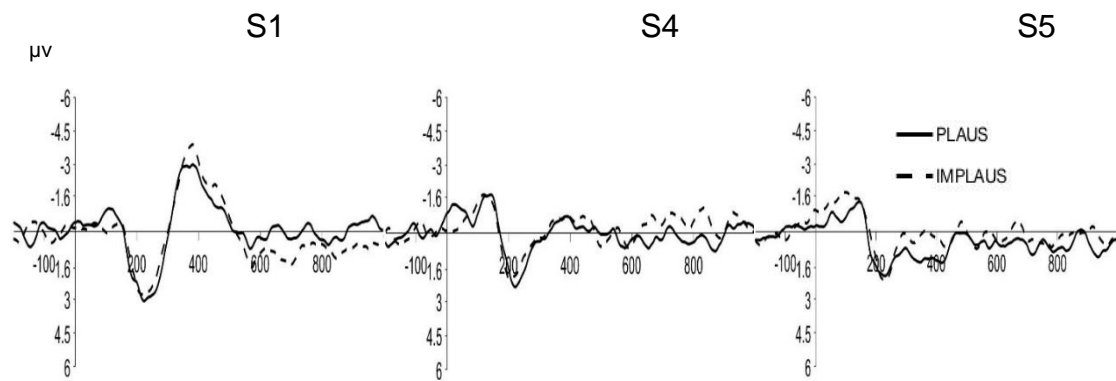


Figure 5. Time course of Event-Related Potentials for the break effects control (S1= Sentence 1, S4= Sentence 4, S5 = Sentence 5). ERPs elicited by target words in semantically plausible sentence stimuli are depicted by the solid line; ERPs elicited

by target words in semantically implausible sentence stimuli are depicted by the dashed lines.

Discussion

In the current experiment, we aimed to address the potential for artefacts created by the break within the Joint Comprehension paradigm. The Joint Comprehension task outlined within Chapter 2 involves the participant reading S1 – S4 alone whilst their partner has closed eyes. The participant then breaks to prompt the partner to open their eyes, and both jointly read S5. In the current experiment, this is instead reversed such that the participant and the partner read S1 – S4 jointly, the participant then breaks to prompt the partner to close their eyes, and the participant reads S5 alone. The results of the current experiment showed a significant N400 on S1, which attenuated such that there is no significant N400 on S4 or S5.

The break during which the partner opens their eyes within the original Joint Comprehension task has potential to introduce both dishabituation and memory interference effects to the experiment. In the reversed paradigm outlined in the current experiment, this break event is maintained, but the joint comprehension aspect of S5 is removed. The participant does not jointly comprehend S5 with the partner, and, if they consider the partner's perspective during S5, the partner will have full context available. Despite maintaining this break to prompt the partner to close their eyes, the planned directional post-

hoc t-tests showed there is no significant N400 on S5. We can therefore infer that the break does not provide a window for dishabituation, and that the attenuation of the N400 between S1 and S4 is not based on habituation of the critical noun, as structurally the current experiment and that in Chapter 3 are matched. We can also infer that the semantic context given in S1 – S4 remains available to the participant to couch interpretation of S5, even after the break.

However, the current experiment suffers from several potential drawbacks. Although the post-hoc t-tests were directional (such that the p -values used are halved) to increase the robustness of the argument for accepting or rejecting the null, the estimated Bayes factor indicates weak support for the null hypothesis that there will be no N400 on S5 (Raftery, 1995). The current study may provide an initial indication but lack statistical power to conclusively accept or reject the null hypothesis. Further, an outcome which relies on the null hypothesis being confirmed could be weaker than one in which we aim to disconfirm the hypothesis.

One potential way of addressing this could be to design a task which reduces the burden of the break on the participant, but still employs joint comprehension on S5. For example, if, in the standard Joint Comprehension task, the participant does not have to remember to remind the confederate to open their eyes, this could mean less opportunity for retroactive interference. Similarly, it would provide less of a window for dishabituation. If the confederate could be cued to open their eyes by a short external factor – such

as a tone being played through headphones to the confederate – this reduces memory and segmentation effects for the participant but allows both participant and confederate to jointly comprehend on S5. Thus, we would predict a significant N400 on S5. The difficulty here is whether the mere change in confederate attention is enough to be considered a boundary event for the participant – as individuals have been shown to segment where the goals or intentions of actors change (Swallow et al., 2009). Playing the tone over headphones to only the confederate might reduce the cues to segment for the participant, although the participant must be cognisant of joint attention on S5.

The results of the current experiment suggest that memory interference effects cannot conclusively be an alternative explanation for the Social N400 seen in Chapter 3. However, it does not yet address whether domain-general non-social cues could explain the current results through submentalising (Conway et al., 2017; Ferguson et al., 2017; Heyes, 2014; Santiesteban et al., 2014, 2015), rather than a reflection of true Theory of Mind, i.e., in the main paradigm, the participant opening their eyes and attending to S5 could be a key attention orienting signal, highlighting S5 as salient to an agent. The next chapter therefore aims to address whether the Social N400 is specifically induced by a human partner or can be induced using non-social agent cues.

Chapter 5 Object-Based Non-Social Cues

Introduction

When determining whether the Social N400 is explained through mentalising (implicit Theory of Mind) or submentalising (domain-general tools), we must determine the components of social cues compared to other types of cues. To claim implicit Theory of Mind, the task must demonstrate reliance on inference about mental states (difficulties in partner comprehension in the absence of context), over and above domain-general signal-dependent goal-directed behaviour (Bowler et al., 2005). In the latter, could it be possible that the confederate is an agent providing a signal (attention to sentence 5) that mediates a goal (reading sentence 5), and that this invokes the effect without necessarily considering the confederate's mental state?

Consideration of what makes a task reliant on social abilities over domain-general abilities has been explored extensively within false belief tasks. Children on the Autism Spectrum struggle to pass the Sally-Anne task more than neurotypical peers (Baron-Cohen et al., 1985) however show higher task performance in a false-photograph analogue, a nonmental-state task (Leekam & Perner, 1991; Perner & Leekam, 2008). This contrasts with results demonstrating that children on the Autism Spectrum show comparable struggles in passing the Sally-Anne task and a mechanical analogue where a

train travels to meet a plane based on a light indicator, and this light indicator can be either true or false (Bowler et al., 2005). In both the Sally-Anne and the Plane-Train task, successful performance requires participants to know that 1) agents act toward goals, 2) action can be mediated by a signal, and that 3) signals can be false. Although the train does not have a mental state, it does represent an agent (*def.* “a person or thing that produces a particular effect or change”, Cambridge Dictionary, n.d.), which may represent the key difficulty difference for Autistic children between this task and the false-photograph task. There are two key highlights here: 1) that some Theory of Mind tasks may be explained outside of mentalisation, instead relying on the understanding that agents act on goals, mediated by signals, and 2) that the involvement of an agent, even without mentalisation, might mediate task performance.

To understand the processes behind the Social N400, we therefore need to establish an appropriate non-mentalising analogue for the Joint Comprehension task. This would need to provide appropriate signal to the participant of “joint attention”, and also provide the conceptualisation of agency in a task in which the “effect” is entirely mental (i.e., the confederate reads silently, and their comprehension of sentence 5 is internal). The study here proposes a camera as an alternative to the confederate. This is on the basis that 1) cameras involve an attention-directing signal (direction of lens) 2) cameras can record a true representation of the event (analogous to passive

reading) but 3) the camera itself cannot hold a mental state regarding the record (there is no conceptualisation of the comprehension of the camera).

Using a camera as a replacement agent, previous research has aimed to distinguish mentalising from submentalising through the use of a Director task (Krauss & Glucksberg, 1977) (see Chapter 1 for details). The study used a human as a Director in one condition and camera as the Director in an inanimate object condition (Santiesteban et al., 2015). If participants were implicitly mentalising they would disregard competitor objects the Director can't see in the human Director condition, but would be equally likely to choose a competitor object in the camera Director condition since the camera does not have a mental state. However, if the participants were using domain-general submentalising tools, their response to the camera Director would be the same as the response to the human Director. The Santiesteban et al study indicated that participants responded to a camera in much the same way a human Director, indicating that some seemingly social impacts may be prompted by domain-general non-social signals.

In the current experiment we replaced the human confederate with a camera in the Joint Comprehension task (Chapter 2). We simulated the 'eyes closed' and 'eyes open' conditions by covering and uncovering the camera lens. The experiment rests on three key assumptions regarding treatment of the camera: 1) that the camera lens provides a signal for attention direction, 2) that the

camera has the capacity to record the event, 3) that the camera does not have a mental state associated with the event. If the effect could be explained by non-social domain-general submentalising we would expect to see effects normally interpreted as the Social N400 despite no human task partner being used. At outset, the premise here is that the Social N400 requires mentalising to consider the comprehension of a partner: thus, we predict no significant N400 on S5.

Methods

The full Joint Comprehension task is outlined in Chapter 2 and is used here with the key modification outlined in 5.1.2.

5.1.1 Participants

A total of 19 participants aged between 18 and 52 years were recruited through University College London (7 men; median age = 25 years, 3 months; SD = 11 years, 3 months). Eight participants were removed from the final analysis due to too few trial contributions following excessive noise in the EEG signal. Therefore, 11 participants remain in the final analysis (5 men; median age = 25 years, 3 months; SD = 11 years, 10 months).

5.1.2 Procedure

The procedure was identical to that outlined in Chapter 2 apart from instruction regarding the covering and uncovering of the camera lens. Prior to sentence 1

instead of seeing the instruction ‘close eyes’ participants saw the instruction ‘cover camera lens’. Participants then covered the camera lens with a lens cap. The lens remained covered until after sentence 4 when participants saw the instruction ‘uncover camera lens’ and removed the cap. In addition, the questions at the end of each trial were changed (from Chapter 2) to Q1 ‘*Do you think the last sentence was plausible for the camera?*’ and Q2 ‘*Was the last sentence plausible for you?*’

5.1.3 EEG Analysis Procedure

5.1.3.1 Processing of EEG data

The average number of trials included in the analysis was S1 PLAUS = 32, S1 IMPLAUS = 30, S4 PLAUS = 32, S4 IMPLAUS = 31, S5 PLAUS = 31, S5 IMPLAUS = 31.

5.1.3.2 Difference in the N400-Effect across Sentence Conditions

As there was no expected change to the N400 on S1 between this experiment and that outlined in Chapter 3, the time window was not recalculated. To directly compare the current results to the Social N400 effect measured in Chapter 3, the mean amplitude of the ERP signal from the midline scalp electrode sites was extracted from each participant for each of the six experimental conditions (PLAUS, IMPLAUS at each of the 3 sentence positions), based on the time window shown to display a robust Canonical N400 on S1 in Chapter 3 (350-400ms). Amplitude of the N400 at each target

time point (S1, S4, S5) was calculated using the difference in amplitude between PLAUS and IMPLAUS, averaged over the midline (Fz, Cz, Pz; in line with Chapter 3). Planned directional (p -value halved) one sample t -tests (Holm's Sequential Bonferroni Procedure used to account for multiple comparisons) were run on mean amplitudes to test for a significant N400 in each sentence position, as in Chapter 3.

5.1.3.3 Bayes Factor

The critical outcome of this study predicts a null result for S5. To further test this, amplitude of the N400 on S5 was entered into a Bayesian One-Sample t -test to contrast the likelihood of the data fitting under the null hypothesis with the likelihood of fitting under the alternative hypothesis.

Results

5.1.4 Behavioural Analysis

Responses to the two questions posed after each trial were analysed in a 2 x 2 repeated measures ANOVA with Question (Q1, Q2) and Condition (PLAUS, IMPLAUS) as main factors. Accuracy rates were as follows Q1-PLAUS: $M = 98.00\%$, $SD = 3.05\%$, Q1-IMPLAUS: $M = 94.01\%$, $SD = 6.49\%$, Q2-PLAUS: $M = 97.34\%$, $SD = 5.05\%$, Q2-IMPLAUS: $M = 93.57\%$, $SD = 7.58\%$, indicating that participants remained engaged in the task. There was no main effect of Question ($F(1, 10) = 0.16$, $p = .698$). There was a main effect of Condition ($F(1, 10) = 7.44$, $p = .021$, $\eta^2 = .43$). There was no significant interaction between

Question and Condition ($F(1, 10) = 0.03, p = .876$). Overall results suggest participants generally found the IMPLAUS condition harder to understand, from both their own and the camera's perspective.

5.1.5 EEG Analysis

5.1.5.1 Differences in the N400-Effect across Sentence Conditions

In order to preserve comparison to the analysis in Chapter 3, mean amplitude of the ERPs elicited by the target words in all six conditions (S1, S4, S5 for both PLAUS and IMPLAUS trials) were calculated in the time window 350-400ms and entered into a 2x3x9 repeated measures ANOVA with the factors Condition (PLAUS, IMPLAUS), Sentence (S1, S4, S5), and Electrode (F3, Fz, F4; C3, Cz, C4; P3, Pz, P4). Unlike Chapter 3 there was no main effect of Condition, $F(1,10) = 0.12, p = .736, \eta p^2 = .01$. There was a main effect of Sentence with Greenhouse-Geisser correction, $F(1.21,12.10) = 7.44, p = .004, \eta p^2 = .43$. There was no main effect of Electrode with Greenhouse-Geisser correction, $F(1.72,17.18) = 1.18, p = .324, \eta p^2 = .11$. The interaction between Condition x Sentence x Electrode was also not significant, $F(16,160) = 1.17, p = .301, \eta p^2 = .10$. Planned comparisons between the plausibility conditions (IMPLAUS < PLAUS) at each sentence position demonstrated a significant effect on S1, $t(10) = -2.87, p = .009$, but not on S4, $t(10) = 1.08, p = .153$ (Holm's Sequential Bonferroni $\alpha = .025$) or S5, $t(10) = .04, p = .484$. An estimated Bayes factor (null/alternative) for N400 amplitude on S5 suggested the data were 4.48:1 in favour of the null hypothesis, or 4.48 times more likely

to occur under a model where mentalising was required compared to one reliant on domain-general cues.

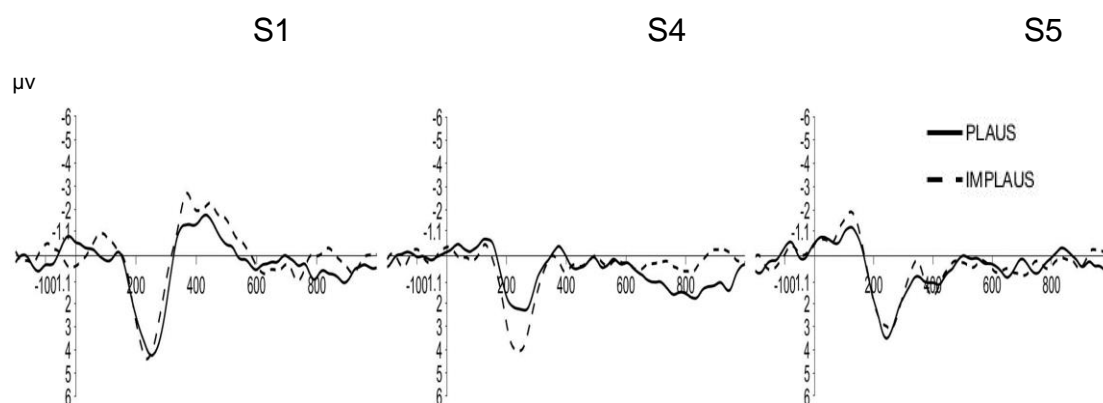


Figure 6. Time courses of Event-Related Potentials for the non-social control (S1= Sentence 1, S4= Sentence 4, S5 = Sentence 5). ERPs elicited by target words in semantically plausible sentence stimuli are depicted by the solid line; ERPs elicited by target words in semantically implausible sentence stimuli are depicted by the dashed lines.

Discussion

In this experiment we adapted the Joint Comprehension task by replacing the human confederate with a camera. Our aim was to test whether the Social N400 effect could be elicited in a non-mentalising analogous task. This would call into question the origins of the effect as something ‘social’ and would raise the possibility that the effect is due to domain-general signal cues that could be elicited by a camera at a critical point in the procedure. Such a finding would be consistent with ‘submentalising’ as described by Heyes (2014). By contrast

we postulate that finding no re-emergence of the N400 between sentence 4 and 5 when a camera is used strengthens the case for the N400 on S5 being 'social'.

The results of the current experiment contrast to previous evidence (Bowler et al., 2005; Santiesteban et al., 2015) that apparent perspective-taking effects can be prompted using non-social cues, and supports the Social N400 as a measure of mentalising. The comparison with these two previous pieces of work could help us to address the concern as to whether the camera was assigned agency within this task. Bowler et al., (2005) hypothesise that a key correlate in the treatment of the Sally-Anne and Plane-Train tasks is the assignment of agency to the acting party. Using this, we could therefore postulate that for Santiesteban et al., (2015) to achieve their results, the camera must have been assigned agency. Since the camera could not act outwardly, this agency appears to be assigned based on the passive act of record: i.e., the camera has the capacity to capture a true record of the other side of the shelves, in much the same way that a human Director is passively viewing the shelves. The camera in our experiment holds the same capacity to capture a true record of sentence 5 to be assigned agency by the participant; it does not, however, have the capacity to manipulate that information within a mental state. Thus, the Social N400 appears to be exclusively cued either by requirement for mentalising about an agent, or the presence of an agent with an internal mental state.

One aspect that could be interesting to modulate here is the extent to which the participants believe the record from the camera will be used by a party with an internal mental state. The step removal could be interesting for building a model of Theory of Mind – the two-system model does not seem to have an account for implicit Theory of Mind where a social agent is not present, whereas a competition model of Theory of Mind might predict that knowledge that the record will be used by an agent could increase the weight of the cue to use implicit Theory of Mind to complete the task in a resource-efficient manner. This would be in line with results from visuo-spatial joint action tasks in which the co-actor does not need to be present, but their goal needs to be conceptualised (Sebanz et al., 2005).

The results of the current experiment may also be surprising, given the study contained several prompts which could encourage participants to anthropomorphise the camera. Participants were asked '*Do you think the last sentence was plausible for the camera?*' a question which assigns linguistic understanding capabilities to the inanimate object. Previous research shows that when individuals anthropomorphise objects, the same areas of the brain that process social cognition are engaged (Gazzola et al., 2007). Despite this, the Social N400 was not invoked by the object. Given the current experiment confounds some elements of both spatial attention-direction and anthropomorphising, it may be useful to further refine the task to strip aspects of the latter, such that the only cue is the attention-direction cue from the

camera lens, although arguably in such a scenario the Social N400 is less likely to be prompted than in the current experiment.

The question of task design is necessarily tied with the critique of the sample size in this experiment. Participants themselves covered and uncovered the camera lens, which introduced significant noise to the EEG data for some participants and led to larger loss of data than expected. As such, the sample size here is very small compared to that seen elsewhere when investigating the N400. The Bayesian analysis of the effect on S5 provides some comfort here, as the Bayes factor of 4.48 suggests positive evidence for rejecting the null (Raftery, 1995). However, future refinement of such a task should anticipate higher loss of data and therefore increase sample size or minimise the need for participant movement when conducting sensitive EEG analysis. One suggestion could be to maintain the Joint Comprehension task with a confederate, but to suppress the ability for the participant to reason about the confederate's goal through increased cognitive load such as articulatory suppression (Lövdén & Johansson, 2003). The confederate would therefore provide a signal for attention, but reasoning regarding agent goal would be suppressed and a Social N400 may not be expected.

It is worth noting that, unlike in the previous chapters, in the current experiment participants performed significantly worse when answering the comprehension questions for the IMPLAUS condition. This indicates that participants found IMPLAUS stories harder to parse correctly than PLAUS stories. However, this

discrepancy holds across both Q1 (regarding the camera comprehension) and Q2 (regarding self-comprehension) and participants demonstrated a robust canonical N400 on sentence 1. These results suggest that the small discrepancy in comprehension between conditions should not impact the N400 across the different sentences, and therefore should not have impacted any potential Social N400.

The results of the current experiment support the assumption that the N400 elicited on sentence 5 in the Joint Comprehension task is indeed a Social N400 and cannot be prompted by an inanimate agent with spatial attention-direction properties but no internal mental state. However, given the experiments thus far have all contained explicit social prompting in the form of the comprehension questions, it is not clear whether the Social N400 represents automatic implicit mentalising. In the next chapter we explore whether the Social N400 is shown even when there are no questions regarding the partner's comprehension, and thus whether the effect can be said to be truly implicit and automatic.

Chapter 6 Automatic Implicit Theory of Mind

Introduction

Thus far we have explored, using the Joint Comprehension task, the potential for individuals to represent the understanding of a context-naïve comprehension partner online. The results from Chapter 4 indicate the Social N400 is not an artefact of memory interference (Heyes, 2014). Chapter 5 also shows the Social N400 also appears to be uniquely social in nature, and is not elicited in response to a surrogate inanimate object replacing the confederate, even when this object has clear focus and attention cues (Heyes, 2014; Santiesteban et al., 2015). Whilst the previous chapters highlight this online simulation of language comprehension, it is not yet clear how this fits into current theories surrounding social perspective-taking, or whether this response is a direct consequence of the directions given within the task. In all previous versions of the task, participants have been explicitly prompted to consider the perspective of their partner, meaning little is known about whether this Social N40 effect can still occur without explicit prompting. Is there an *implicit* Social N400 operating in the context of language comprehension?

The current two-factor model of Theory of Mind (Apperly & Butterfill, 2009; Butterfill & Apperly, 2013; Frith & Frith, 2008) hypothesises that the implicit system is both fast and automatic. This has some support from evidence the system is present from infancy (Southgate et al., 2007; Southgate & Vernetti,

2014), appears to be based on involuntary embodiment of the perspectives of others – from social to visual (Hasson & Frith, 2016; Schneider, Bayliss, et al., 2012), and is evidenced within fast and unconscious looking patterns (Schneider et al., 2013; Senju, 2012, 2013; Senju et al., 2009).

Despite this evidence for an apparent two-stage Theory of Mind, there remains debate surrounding the neural and cognitive basis for an implicit Theory of Mind. One theory postulates that implicit Theory of Mind could be based within embodiment of others we interact with. Within action perception research, “mirror” neurons have been found that fire both when we complete an action, and when we see another completing the same action – suggesting interpretation of the actions of others may be based within embodying those actions within our own action systems (Gallese, 2007; Gallese & Goldman, 1998; Rizzolatti & Craighero, 2004). This embodiment is involuntary and happens without prompting. Behavioural correlates for embodiment of others have been shown across joint action (Vesper et al., 2017), visual perception (Samson et al., 2010), and false-belief tasks (Senju, 2012; Southgate et al., 2007). The Social N400 effect shown in Chapter 3 appears to build into this growing literature showing the mental states of others are reflected on-line within our own cognitive systems. However, the current task fails to assess whether this apparent embodiment of comprehension used in joint comprehension is spontaneous, as suggested within the action perception research.

A growing body of literature also casts doubt on the existence of a truly implicit mentalising system for Theory of Mind, independent of the higher-order executive function abilities underpinning explicit Theory of Mind. For example, eye-movements thought to be indicative of implicit Theory of Mind can be disrupted whilst the participant is under additional cognitive load (Schneider, Lam, et al., 2012; Scott & Roby, 2015) suggesting at least minimal recruitment of executive functions in these tasks. fMRI data also suggests that implicit and explicit Theory of Mind share neural correlates (Van Overwalle & Vandekerckhove, 2013) and explicit tasks may simply demand increased resources from the same areas (Lewis et al., 2017). So-called “automatic” Theory of Mind processes have also been shown to be modulated by motivation and reward (Cane et al., 2017), and anticipatory looking tasks fail to be replicated consistently (Burnside et al., 2018; Kulke et al., 2018). This therefore casts doubt on the existence of two independent social systems and raises the question: if both implicit and explicit Theory of Mind are reliant on similar mechanisms, why is one automatic and the other effortful?

The full body of evidence described could suggest that whilst explicit and implicit Theory of Mind processes are separate tools, they may not represent separate temporally bound systems. Such a model might instead suggest implicit embodiment may be a mechanism capable of being recruited within Theory of Mind, in much the same way as suggested within the competition model of language. Where cues are weighted appropriately to be used for the

task, they will be recruited (MacWhinney, 1992; Siegal & Varley, 2002). Where they do not meet the threshold for resource use, they will be discarded. This therefore raises the question: is our early Social N400 automatic and always employed in the presence of a co-listener, or is there a level of cue weighting needed before online modelling is used?

To date, the Joint Comprehension task involved asking participants after each trial “Do you think the last sentence was plausible for your partner?” thereby building expectation that social inference would be needed, and potentially priming social perspective-taking systems to be employed. In the current study we remove the explicit questioning about the confederate’s understanding, and instead move the task to be solely focussed on the individual’s comprehension. If the Social N400 effect is shown despite reasoning about the confederate’s perspective no longer being explicitly necessary for the task, this would support the hypothesis of spontaneous implicit embodiment of others. If the Social N400 is not shown in this new scenario, it would suggest that embodiment is a cognitive mechanism that can be employed when social perspective-taking is needed but is not spontaneous.

Methods

The full Joint Comprehension task is outlined in Chapter 2 and is used here with the absence of Q1: ‘Do you think the last sentence was plausible for your partner?’

6.1.1 Participants

A total of 24 participants aged between 19 and 56 years were recruited through University College London (10 men; median age = 24 years, 7 months; SD = 12 years, 11 months). Nine participants were removed from the final analysis due to too few trial contributions following excessive noise in the EEG signal. Therefore, 15 participants remain in the final analysis (6 men; median age = 25 years, 3 months; SD = 15 years, 2 months).

6.1.2 EEG Analysis Procedure

6.1.2.1 Processing of EEG data

The average number of trials included in the analysis was S1 PLAUS = 33, S1 IMPLAUS = 33, S4 PLAUS = 34, S4 IMPLAUS = 35, S5 PLAUS = 34 S5 IMPLAUS = 33.

6.1.2.2 Difference in the N400-Effect across Sentence Conditions

As there was no expected change to the N400 on S1 between this experiment and that outlined in Chapter 3, the time window was not recalculated. To allow comparison between the current results to the Social N400 effect measured in Chapter 3, the mean amplitude of the ERP signal from the midline scalp electrode sites was extracted from each participant for each of the six experimental conditions (PLAUS, IMPLAUS at each of the 3 sentence positions), based on the time window shown to display a robust Canonical N400 on S1 in Chapter 3 (350-400ms). Amplitude of the N400 at each target

time point (S1, S4, S5) was calculated using the difference in amplitude between PLAUS and IMPLAUS, averaged over the midline (Fz, Cz, Pz; in line with Chapter 3). Planned one sample t-tests (directional for S1 and S4, bi-directional for S5; Holm's Sequential Bonferroni Procedure used to account for multiple comparisons) were run on mean amplitudes to test for a significant N400 in each sentence position, as in Chapter 3.

6.1.2.3 Bayes Factor

In the absence of a strong directional hypothesis for the N400 effect on S5, to further test any effect, amplitude of the N400 on S5 was entered into a Bayesian One-Sample t-test to contrast the likelihood of the data fitting under the null hypothesis with the likelihood of fitting under the alternative hypothesis.

Results

6.1.3 Behavioural Analysis

Responses to the question posed after each trial were analysed using within subjects t-test comparing accuracy between PLAUS and IMPLAUS. Accuracy rates were high (PLAUS: $M = 96.91\%$, $SD = 3.96\%$, IMPLAUS: $M = 96.75\%$, $SD = 3.41\%$), indicating that participants remained engaged in the task. There was no significant effect of Condition, $t(1, 14) = 0.24$, $p = .818$, indicating that both conditions had similar levels of correct interpretation.

6.1.4 EEG Analysis

6.1.4.1 Differences in the N400-Effect across Sentence Conditions

In order to preserve comparison to the analysis in Chapter 3, mean amplitude of the ERPs elicited by the target words in all six conditions (S1, S4, S5 for both PLAUS and IMPLAUS trials) were calculated in the time window 350-400ms and entered into a 2x3x9 repeated measures ANOVA with the factors Condition (PLAUS, IMPLAUS), Sentence (S1, S4, S5), and Electrode (F3, Fz, F4; C3, Cz, C4; P3, Pz, P4). There was a significant main effect of Condition, $F(1,14) = 5.27$, $p = .038$, $\eta^2 = .27$ and Sentence, $F(2,28) = 8.67$, $p = .001$, $\eta^2 = .38$. The interaction between Condition x Sentence x Electrode was not significant, $F(16,224) = 0.68$, $p = .817$, $\eta^2 = .05$. Mean amplitude of the ERPs elicited by the target words in all six conditions (S1, S4, S5 for both PLAUS and IMPLAUS trials) were calculated in the time window defined by the S1 N400-Effect (350-400ms). Planned comparisons between the plausibility conditions (IMPLAUS < PLAUS) at each sentence position demonstrated a significant effect on S1, $t(14)=1.98$, $p = .034$, but not on S4, $t(14)=1.21$, $p = .123$, or S5, $t(14) < 1$, $p = .792$. An estimated Bayes factor (null/alternative) for N400 amplitude on S5 suggested the data were 4.97:1 in favour of the null hypothesis, or 4.97 times more likely to occur under a model where the Social N400 was not spontaneous compared to one where it is.

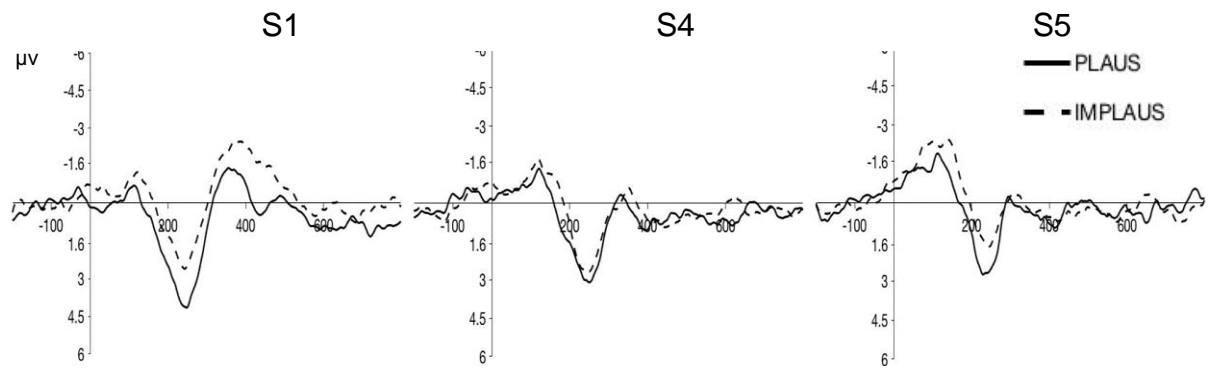


Figure 7. Time courses for Event-Related Potentials for the Automatic Implicit Social N400 (S1= Sentence 1, S4= Sentence 4, S5 = Sentence 5). ERPs elicited by target words in semantically plausible sentence stimuli are depicted by the solid line; ERPs elicited by target words in semantically implausible sentence stimuli are depicted by the dashed lines.

Discussion

In the current study, the Joint Comprehension task was modified to remove explicit questioning about the confederate's semantic understanding. Using this design, it was shown that the Social N400 is not displayed in participants when explicit social prompting is not used. Participants showed the expected attenuation of the N400 with increased semantic context but did not show resurgence of the N400 when the context-naïve confederate was jointly attending to the final sentence of the stories, as was displayed in Chapter 3.

The current results do not support an automatic implicit Social N400 as would be predicted by a two-factor model of Theory of Mind. This model predicts

implicit Theory of Mind is spontaneous where a social partner is present (Apperly & Butterfill, 2009). Instead, the absence of the Social N400 in the absence of prompting to consider a partner's perspective may suggest such an embodiment cognitive mechanism can be employed when the task requires, i.e., when the cue to use implicit mechanisms reaches a threshold for use. Implicit Theory of Mind could still be said to be fast paced (sub 400ms) and based in representing others within the self (representing the comprehension of others within our own language comprehension mechanisms), but instead could be conceptualised as a tool that can be dynamically and flexibly used, as opposed to a rigid dedicated system which is always in use.

Indeed the current results fit in with previous research which has shown that "Implicit" processes are modulated by motivation and reward (Cane et al., 2017) and may further shed light on the non-replication of anticipatory looking tasks (Burnside et al., 2018; Kulke et al., 2018). In this version of the experiment, there is little motivation for participants to incorporate the perspective of their partner, like in the Cane et al (2017) modulation of the Director Task where participants were not rewarded for speed or accuracy of disambiguation. This lends an interesting question to research using anticipatory looking tasks; are the differences between presence and absence of anticipatory looking modulated by the perceived motivation of the task, rather than simple presence or absence of embodied processing?

It would also be interesting to explore the extent to which a social partner needs to be involved before an implicit Social N400 might be shown. In the current experiment, it is relatively easy for the participants to ignore the confederate, given their presence has no relevance to the task. If the confederate was perhaps joined in the task in some way, or in small ways drawing attention to themselves, might the Social N400 be shown despite no explicit questioning about the confederate's perspective? This could also be considered within the context of the necessity of agent action for a cue to be considered social (Bowler et al., 2005). In the Joint Comprehension task, the prompt regarding the partner's comprehension reminds the participants that the confederate is undertaking an action – they are comprehending the sentence for themselves. In the absence of this prompt, and with no further need to consider the confederate, the confederate's presence could be said to be non-social: there is no need to consider the action of the agent. This could also explain why in other experiments the Social N400 is present in the absence of prompting - the confederate was undertaking a task indicated by a button-press and so could have been more prominent for the participants (Jouravlev et al., 2019).

As is the case with much of the evidence presented here, the sample size for this experiment is small, although not unusual for EEG investigations into the N400 (Bhavnani et al., 2021). The Bayesian analysis of the effect on S5 provides support for rejecting the null, as the Bayes factor of 4.48 suggests

positive evidence of the data fitting a model in which the Social N400 is not present in the absence of task prompting (Raftery, 1995).

Whilst the previous experiments in Chapter 4 and 5 showed that the Social N400 can be said to be truly social in nature, and not explained by domain-general submentalising, the effect also does not appear to reflect true automatic implicit Theory of Mind as currently hypothesised within the two-system model. The recruitment of participants' own language comprehension systems to reason about the confederate's understanding aligns with the theory that aspects of social perspective-taking reside within embodiment within one's own systems and suggests instead that this may be a tool employed when it would aid with a task. Rather than a separate system, implicit Theory of Mind therefore becomes a tool, and Theory of Mind achieved through the coordination of multiple cognitive tools.

Chapter 7 The Social N400 in Adolescents

Introduction

Thus far through this body of work, we have examined the use of a new Joint Comprehension task to establish the existence of a Social N400, a potential correlate of implicit Theory of Mind. Within Theory of Mind research, there is plenty of evidence that implicit Theory of Mind, either as a system or as a cognitive tool, is available in pre-adult populations, potentially even from infancy (Southgate et al., 2007; Southgate & Vernetti, 2014). To state that the Social N400 is indeed indicative of implicit Theory of Mind, it must therefore be established pre-adulthood. Given the current task is heavily reliant on developed language skills, this chapter looks to investigate the Social N400 in an adolescent population, using the Joint Comprehension task.

Unlike in younger children, adolescent linguistic capability is on par with that of adults (Cummings et al., 2008; Hahne et al., 2004; Holcomb et al., 1992; Juottonen et al., 1996; Lidzba et al., 2011). This serves to increase confidence that, when using the Joint Comprehension task, adolescents will be sensitive to changes in semantic congruency for themselves. The N400 as a marker of semantic incongruency has also been shown to be available in adolescents, although the marker may differ in size, latency, and location to its equivalent in adults (Holcomb et al., 1992). Whilst this may impact the way in which we

identify the N400 within the analysis, overall, we would expect the N400 to behave similarly in adolescent to in adults.

In addition, although adolescents show improvement in performance in perspective-taking tasks into adulthood (Burnett et al., 2008; Dumontheil et al., 2010; Pfeifer & Blakemore, 2012), implicit Theory of Mind has been demonstrated behaviourally from a young age (Blakemore & Choudhury, 2006; Nadig & Sedivy, 2002; Scott & Baillargeon, 2017). In an experiment with participants as young as four, children were shown to identify ambiguous targets quicker when adults built up shared references over time and were delayed when these references were violated, for example moving from describing a dog as “the spotted dog” to “the fluffy dog” despite both adjectives being true (Graham et al., 2014). This effect may not be partner-specific, indicating that children may not appreciate the context built up with one speaker may not generalise to another speaker (Matthews et al., 2010), or indeed might indicate that children have a lower threshold at which implicit Theory of Mind is used compared to adults. This would correlate with the importance of social cues within adolescent life, whilst individuals are still learning about the world around them and their role within it (Blakemore, 2018a, 2018b).

The current study tests the within-subjects Joint Comprehension task with adolescent participants. As adolescents show adult-level language capabilities, we hypothesise that participants will display an N400 on S1 based

on animacy violations, which will be attenuated by S4 such that there is no significant N400. Based on our previous experiments, the Social N400 appears to be a candidate for a marker of implicit Theory of Mind within a competition model – a cognitive tool for completing Theory of Mind tasks where the task demands it. Given the evidence that implicit Theory of Mind is present prior to adulthood, we hypothesise adolescents will show a Social N400 on S5, indicating adolescent participants are representing the comprehension of a joint reading partner within their own cognitive mechanisms for comprehension.

Methods

The full Joint Comprehension task is outlined in Chapter 2 and is used here. Data for this experiment was collected at the Department of Psychology, University of York, and therefore the EEG analysis is outlined separately in 7.1.2 due to differing EEG recording equipment and differences in processing adolescent data.

7.1.1 Participants

A total of 23 adolescents aged between 10 and 15 years were recruited through local schools (10 boys, mean age = 12 years, 5 months, SD = 1 year, 7 months). Seven participants were removed from the final analysis: one due to a technical error during recording, and six due to too few trial contributions following excessive noise in the EEG signal (< 10 trials remaining), or poor

comprehension of the task (< 75% correct). Therefore, 16 adolescents remain in the final analysis (mean age = 12 years, 6 months; SD = 1 year, 7 months).

7.1.2 Stimuli

As stated in Chapter 2, catch trials were not included in the current experiment to simplify the task for the younger population.

7.1.3 EEG Analysis Procedure

7.1.3.1 Processing of EEG data

EEG data processing differs here from the previous procedure outlined in adult participants. This is due to a combination of the differing EEG recording systems used (e.g., sampling rate differences) and increased noise experienced within adolescent participant data due to the testing environment available (addressed using stricter bandpass filter parameters to better focus on activity in theta and delta frequencies related to the N400 (Steele et al., 2013)).

Continuous EEG was recorded using ASALab in a quiet room from 32 shielded active electrodes placed in a 10-20 montage (recording reference = M1, ground = forehead, VEOG and HEOG included, electrode impedances < 10 k Ω , bandpass filter 0.5-100Hz, notch filter = 50Hz, sampling rate = 500 Hz, resampled to 200 Hz off-line). Data analysis and pre-processing were conducted using EEGLab and ERPLab. Data were re-referenced off-line to the

average of the mastoids and filtered using a 0.1 – 20 Hz bandpass filter. Continuous EEG signal was visually inspected for major artefacts, and individual channels with excessive noise were interpolated as an average between two nearest neighbours (Planner interpolation). The signal elicited by correctly answered trials was segmented (-200 to 1000ms relative to the onset of the target word in each sentence) and a semi-automatic artefact rejection using a 100ms Moving Window, amplitude spikes > 100 μ v as well as visual inspection were applied to reject segments with excessive noise. The average number of trials included in the analysis was S1 PLAUS = 29, S1 IMPLAUS = 27, S4 PLAUS = 30, S4 IMPLAUS = 29, S5 PLAUS = 26, S5 IMPLAUS = 24. Baseline correction was applied to the time window -200ms to 0ms relative to the onset of the target word. Pre-processed segments were then averaged per condition within participants.

7.1.3.2 Canonical N400-Effect

Prior evidence indicates that the N400 in an adolescent population may differ in both location and latency when compared to adults. Whereas in the adult population the location was known and therefore the time window was calculated using tmax permutation analysis, here we use a two-tailed cluster mass permutation test to identify both location and time window of the N400.

To determine the time window in which reliable differences between PLAUS and IMPLAUS stimuli were seen on S1 (i.e., a canonical N400-Effect), the ERP signal from these conditions were submitted to a repeated measures, two-

tailed cluster mass permutation test using a family-wise alpha level of 0.05 (Bullmore et al., 1999; Groppe et al., 2011) using the Mass Univariate ERP Toolbox. All time points between 100 and 900ms post presentation of the critical word in each condition on S1 at all 30 scalp electrode sites were included in the test, and any electrodes within approximately 5.44 cm of each other were considered spatial neighbours. Repeated measures t-tests were performed for each comparison using the original data and 10,000 random within-participant permutations of the data. For each permutation, all t-scores corresponding to uncorrected p-values of 0.01 were formed into clusters. The sum of the t-scores in each cluster defines the mass of the cluster, and the most extreme cluster mass in each of the 10,001 sets of tests was recorded and used to estimate the time window and distribution of the null hypothesis.

7.1.3.3 Difference in the N400-Effect across Sentence Conditions

To compare the N400 across sentence conditions, the mean amplitude of the ERP signal from three central electrodes that showed a robust difference across the N400 time window in the cluster permutation analysis (C3, Cz, C4) was extracted from each participant for each of the six experimental conditions (PLAUS, IMPLAUS at each of the 3 sentence positions). To assess whether significant differences between the conditions could be seen across sentence positions, mean amplitude changes were first entered into a 2x3 repeated measures ANOVA with the factors Condition (PLAUS, IMPLAUS) and Sentence (S1, S4, S5). Planned directional (p -value halved) one sample t-tests

(Bonferroni Correction used to account for multiple comparisons) were run to test for differences between conditions in each sentence position.

7.1.3.4 Latency of N400-Effect

In addition, the peak latency of the N400-Effect elicited on S1 and S5 was extracted and entered into a paired samples t-test in order to test for differences in the latency of the effect across sentence positions.

Results

7.1.4 Behavioural Analysis

Responses to the two questions posed after each trial were analysed in a 2 x 2 repeated measures ANOVA with Question (Q1, Q2) and Condition (PLAUS, IMPLAUS) as main factors. Accuracy rates were generally high (Q1-PLAUS: M = 86.90%, SD = 9.08%, Q1-IMPLAUS: M=87.38%, SD = 9.60%, Q2-PLAUS: M = 96.39%, SD = 4.49%, Q2-IMPLAUS: M = 90.50 %, SD = 11.93%), indicating that participants remained engaged in the task. A main effect of Question was observed: more errors were made when participants were asked about the confederate's perspective (Q1) vs. their own perspective (Q2), $F(1, 15) = 14.16$, $p < .005$, $\eta^2 = .49$. No main effect of Condition was observed, $F(1, 15) = 1.39$, $p > .10$, indicating that neither Condition was significantly more difficult for participants than the other. In addition, a significant interaction between Question x Condition was observed, $F(1, 15) = 10.15$, $p < .01$, $\eta^2 = .40$. Post-hoc paired samples t-tests were run to resolve the interaction; the

Bonferroni corrected significance threshold was set at $\alpha = 0.0125$ to account for multiple post-hoc comparisons. This analysis revealed a trend in participants' performance rates when judging IMPLAUS sentences vs. PLAUS sentences when asked about their own perspective (Q2), $t(15) = 2.76$, $p = .015$, but no similar difference between the two sentence conditions when asked about the confederate's understanding (Q1), $t(15) = 0.17$, $p > 0.1$. The larger number of errors made for one's own interpretation of IMPLAUS stimuli suggests that not all implausible scenarios were successfully mitigated by discourse context; where discourse context is irrelevant (i.e., in answering Q1), no difference in the number of errors is seen between conditions.

7.1.5 EEG Analysis

7.1.5.1 Canonical N400-Effect

The cluster mass permutation analysis revealed just one significant cluster ($p < 0.001$) broadly distributed across centro-parietal electrodes over both hemispheres in the time window from 365-630ms ($p < 0.001$). In this time window, IMPLAUS target words elicited a significantly stronger negative signal than PLAUS target words. The time window and topography of this effect is consistent with those generally seen for the N400-Effect (see Figure 8).

7.1.5.2 Differences in the N400-Effect across Sentence Conditions

Mean amplitude of the ERPs elicited by the target words in all six conditions (S1, S4, S5 for both PLAUS and IMPLAUS trials) were calculated in the time

window defined by the S1 N400-Effect (365-630ms) and entered into a 2x3 repeated measures ANOVA with the factors Condition (PLAUS, IMPLAUS) and Sentence (S1, S4, S5). Main effects of both Condition, $F(1,15) = 10.75$, $p = .005$, $\eta^2 = .42$, and Sentence, $F(2,30)=18.42$, $p < 0.001$, $\eta^2 = .55$, were observed. In addition a significant Condition x Sentence interaction was observed, $F(2,30) = 3.42$, $p < .05$, $\eta^2 = .19$, indicating that plausibility affected the signal differently at the different sentence positions. Planned comparisons between the plausibility conditions (IMPLAUS < PLAUS) at each sentence position demonstrated a significant effect on S1, $t(15) = 4.81$, $p < .001$ and S5, $t(15) = 2.32$, $p < .016$ (Bonferroni $\alpha = .016$), but not on S4, $t(15) < 1$, $p = .450$).

7.1.5.3 Latency of the N400-Effect

There was no difference between the peak latency of the N400-Effect elicited on S1 (Mean = 508ms, SD = 95ms) and S5 (Mean = 494ms, SD = 92ms), $t(15) = .42$, $p > 0.1$.

Figure removed due to copyright.

For original figure, please see <https://doi.org/10.1016/j.jecp.2017.06.016> (Figure 2)

Figure 8. Event-related potentials and scalp topography for the Social N400 in adolescents. Time courses of ERPs can be seen for target words in each sentence position (S1= Sentence 1, S4= Sentence 4, S5 = Sentence 5). ERPs elicited by target words in semantically plausible sentence stimuli are depicted by the solid line; ERPs elicited by target words in semantically implausible sentence stimuli are depicted by the dashed lines. In the bottom panel scalp distributions showing differences between the conditions across the N400 time window are seen for each sentence position.

Discussion

The current study aimed to use the Joint Comprehension task to probe the Social N400 in an adolescent population, as would be predicted by behavioural evidence of implicit Theory of Mind. Our results support the assertion of adult-like language abilities in adolescents: adolescents integrate semantic information and make use of additional context in language comprehension; adolescent participants showed a clear N400 effect (a marker of semantic incongruence) on sentence 1 (S1), which then attenuated with additional context repetition, until no N400 marker is shown by sentence 4 (S4). Crucially, our results also suggest adolescents simulated the comprehension of a context-naïve jointly attending confederate for sentence 5 (S5), with participants displaying a Social N400 effect despite parsing sentences correctly for themselves.

Before using the Social N400 to probe implicit mentalising and integrate this into a model of Theory of Mind, the Social 400 must first correlate with results

from behavioural data, e.g., that implicit Theory of Mind is present pre-adulthood. Previous behavioural investigations using implicit measures (e.g., eye tracking, looking patterns) have shown that children and young adolescents can be sensitive to perspectives of others within dialogue. Children as young as four can use linguistic common ground and shared references to narrow down ambiguous choices (Graham et al., 2014; Matthews et al., 2010) and by age six they show looking patterns indicating they can take into account a speaker's visual perspective (Nadig & Sedivy, 2002). The ability for such an implicit mechanism to combine with reasoning-based explicit mentalising appears to still be developing in children (Nadig & Sedivy, 2002) and continues to develop well into young adulthood (Dumontheil et al., 2010), largely due to immature executive function capabilities in adolescents. The results from the current study therefore correlate with previous demonstrations that implicit mentalising, representing another within the self, can be shown in pre-adult populations.

The investigation here centres on whether adolescents display a Social N400, in the same way that they display other behavioural measures but does not seek to compare the Social N400 in adolescents to that seen in adults. Small differences in the procedure might limit such a direct comparison within the current data, including the exclusion of catch trials from the adolescent experiment to simplify the procedure for the young participants. An interesting future investigation could be to directly compare both populations, or indeed

multiple age populations, and correlate this to changes in Theory of Mind success over ages. In doing so, future research may wish to address the sample size here. Although the current sample provides statistically significant results, to aid with future validity and comparability it would be prudent to conduct power analysis using these parameters as a pilot to determine future sample sizes.

A further investigation into how cue usage is modulated in adolescence could greatly inform any future model of Theory of Mind. Within the behavioural data, adolescents appear to use implicit Theory of Mind less sparingly than adults (Matthews et al., 2010). If this holds, a replication of the Joint Comprehension task without prompting to consider the confederate perspective, as in adults in Chapter 6, might be expected to show that adolescents still display the Social N400 without prompting. It could be that the demand to use social information to learn about the world, and to learn to belong in a social world, itself presents a higher weighted cue within adolescence than it does within adulthood where learning might be more limited, and the brain has largely undergone its developmental changes (Blakemore & Choudhury, 2006).

Overall, the current experiment shows that the Social N400 is developed in pre-adult populations and can be used where representing others within the self may aid with task demands. This is a critical qualifier for the hypothesis that the Social N400 represents implicit Theory of Mind, which has previously

been demonstrated in pre-adult populations within extensive behavioural examination.

Chapter 8 Autism Spectrum Disorders: Single Case Study

Introduction

The penultimate chapter of this thesis reports on a pilot study designed to test the feasibility of using the Joint Comprehension task with an individual with an ASD diagnosis. This was with a view to future experiments examining whether Autistic individuals show a Social N400 effect. Autism Spectrum Disorder (ASD) is a neurodevelopmental condition characterised by impairments in two core areas: i) social communication and social interaction, and ii) repetitive behaviours and restricted interests (American Psychiatric Association, DSM-V, 2013). There is considerable evidence that Autistic individuals have difficulty mentalising, tending to be delayed relative to age and IQ matched controls on explicit Theory of Mind tasks (Baron-Cohen, 2000). However, as verbal mental age increases Autistic individuals are more likely to be able to pass explicit false belief tasks (at least when they are first order tasks) leading to the suggestion that they can devise a solution to tasks through explicit reasoning (Happé & Frith, 2014; Happé, 1994). Despite this, those individuals on the Autism Spectrum who can pass explicit false belief tasks do not appear to show evidence of an implicit Theory of Mind (Senju et al., 2010).

Our results thus far require a change to the conceptualisation of Theory of Mind, and the way in which success in Theory of Mind tasks might be achieved for Autistic individuals. Whereas in a two-system model of Theory of Mind

individuals on the Autism Spectrum might be hypothesised to have an 'impaired' implicit Theory of Mind system, with an 'intact' explicit Theory of Mind system (Apperly & Butterfill, 2009; Senju, 2013; Senju et al., 2009, 2010), a competition model of Theory of Mind might hypothesise a different pattern of mechanism recruitment for Autistic individuals to neurotypical individuals. If we conceptualise implicit Theory of Mind as a cognitive tool, this raises several questions about the way we have interpreted evidence from behavioural implicit Theory of Mind tasks to date. Is an implicit cognitive tool available to those on the Autism Spectrum? Is there a differing threshold for implicit Theory of Mind engagement for those on the Autism Spectrum compared to neurotypical individuals? Is implicit Theory of Mind an essential component of the flexible use of social perspective-taking which can be more difficult for those on the Autism Spectrum to achieve?

To begin to address these questions we must first as a baseline establish the presence or absence of the Social N400 using the current Joint Comprehension task in an Autistic population. In the same way that for the Social N400 to represent implicit Theory of Mind it needed to be present in pre-adult populations, for the Social N400 to correlate with current behavioural evidence from implicit Theory of Mind, we should anticipate the effect being absent in Autistic populations.

The current task presents a challenge for exploration with Autistic individuals as it is heavily reliant on language abilities. Difficulties with language

comprehension have been found to be more prevalent in Autistic individuals (Loucas et al., 2008). When compared to neurotypical individuals, performance on language comprehension tasks appears to be modulated by Theory of Mind in Autistic individuals but not neurotypical individuals (McIntyre et al., 2018). A heavy reliance on language comprehension to assess Theory of Mind may therefore introduce significant confounds and limit ability to interpret results.

In this pilot study we were simply concerned with the practical implications of running the task with an Autistic individual, focussing on the outcomes from explicit testing, and the reported experience of our participant. Would the environment of being in a small experimental room next to a confederate be acceptable? Would the instructions make sense, or would we need to make adaptations? Would the behavioural data at least suggest a general comprehension of the task demands, particularly given the reliance on language?

Methods

The full Joint Comprehension task is outlined in Chapter 2 and is used here.

8.1.1 Participant

A single participant was recruited through University College London. Participant I was female; age 50; self-reported a prior clinical diagnosis of ASD

and no further prior history of neurological impairments; a native speaker of English; with corrected-to-normal vision. Participant I demonstrated a Verbal Comprehension score of 87 using the Wechsler Adult Intelligence Scale IV, which is low average – average (Wechsler, 2008).

8.1.2 Stimuli

As stated in Chapter 2, catch trials were not included in the current experiment to simplify the task.

8.1.3 EEG Analysis Procedure

8.1.3.1 Processing of EEG data

The number of trials included in the analysis was S1 PLAUS = 43, S1 IMPLAUS = 44, S4 PLAUS = 48, S4 IMPLAUS = 38, S5 PLAUS = 40, S5 IMPLAUS = 43.

8.1.3.2 Difference in the N400-Effect across Sentence Conditions

To directly compare the current results to the Social N400 effect measured in Chapter 3, the time window for analysis was not recalculated. The mean amplitude of the ERP signal from the midline scalp electrode sites was extracted for each of the six experimental conditions (PLAUS, IMPLAUS at each of the 3 sentence positions), based on the time window shown to display a robust Canonical N400 on S1 in Chapter 3 (350-400ms). Amplitude of the N400 at each target time point (S1, S4, S5) was calculated using the difference

in amplitude between PLAUS and IMPLAUS, averaged over the midline (Fz, Cz, Pz).

Results

8.1.4 Behavioural Analysis

To assess task comprehension, the answers to Q1 *'Do you think the last sentence was plausible for your partner?'* and Q2 *'Was the last sentence plausible for you?'* were outputted for both PLAUS and IMPLAUS conditions. For PLAUS, the participant gave correct answers for both Q1 and Q2 in 100% of trials. For IMPLAUS, accuracy for Q1 was 98.1% and Q2 was 92.3%.

8.1.5 EEG Analysis

8.1.5.1 Differences in the N400-Effect across Sentence Conditions

The graph below shows the ERPs recorded across S1, S4, and S5 within the time period -200ms – 1000ms. Within the critical time period of 350 – 400ms, the amplitude of the N400 response was 1.18 at S1, -0.91 at S4, and -3.59 at S5.

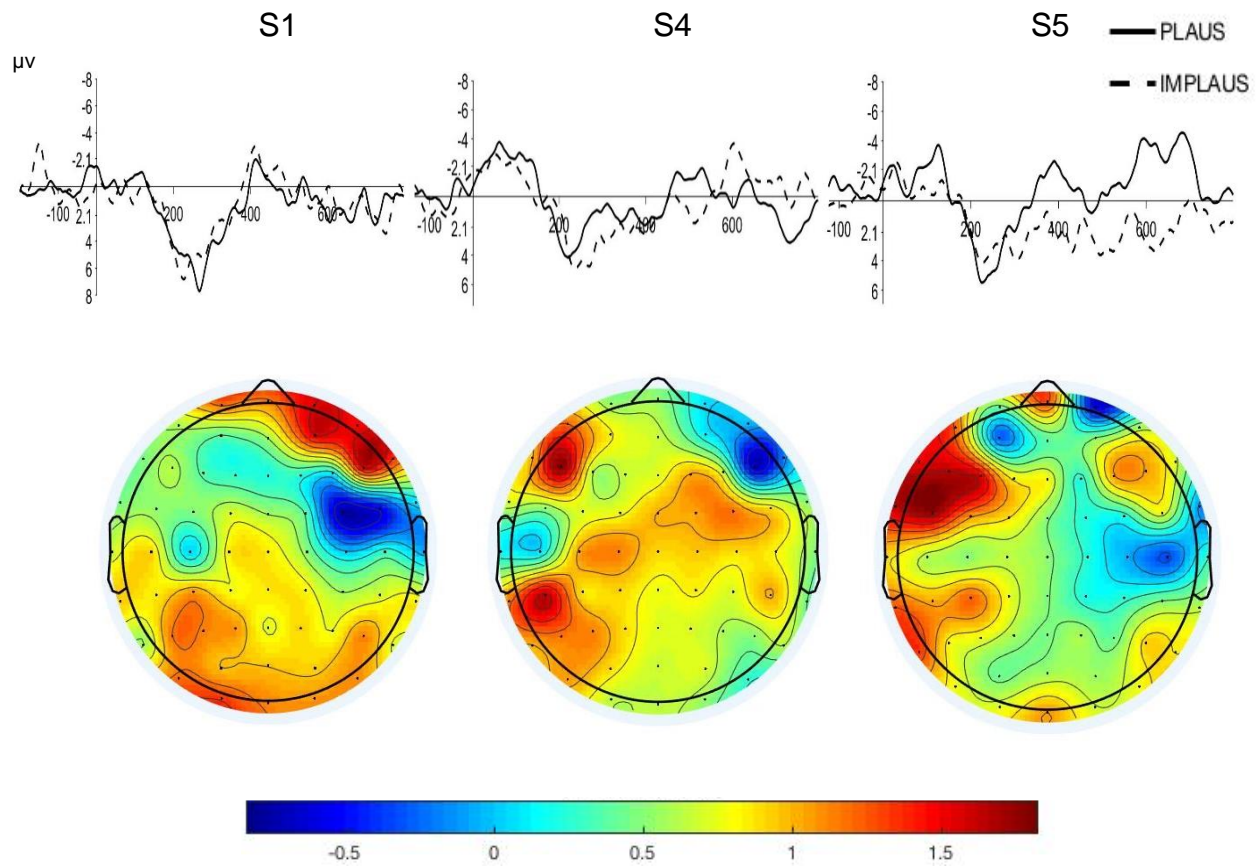


Figure 9. Autism Spectrum Disorder case study: Event-Related Potentials and scalp topography for the Social N400. Time courses of ERPs can be seen for target words in each sentence position (S1= Sentence 1, S4= Sentence 4, S5 = Sentence 5). ERPs elicited by target words in semantically plausible sentence stimuli are depicted by the solid line; ERPs elicited by target words in semantically implausible sentence stimuli are depicted by the dashed lines. In the bottom panel scalp distributions showing

differences between the conditions across the N400 time window are seen for each sentence position.

8.1.6 Task Feasibility

Following the experiment, we conducted a debrief with Participant I. She expressed some surprise at the partner being a confederate, noting that she had been focussed on her own task. She did not think the confederate's role in the experiment was tedious – a potential future modification would be to give the confederate a task also related to sentence 5 so that they are seen to have a purpose. Participant I expressed mild discomfort at the EEG set up, in particular the gel used for the electrodes. The standard setup uses 64 electrodes, however, since many of them are ultimately not used in the final analysis, it may be preferable to reduce the number of electrodes that are filled with gel to minimise discomfort and reduce the length of the experiment.

Generally, Participant I understood the task and correctly answered the questions. She performed worse on Q2 – regarding the self. It could be that Participant I deduced a rule for Q1 – the confederate's comprehension – and then considered her own comprehension without a rule. In future experiments, inclusion of catch trials could help to ensure the participants aren't following a rules-based pattern.

Discussion

The aim of this single case study was to test the feasibility of using the Joint Comprehension paradigm with individuals with a diagnosis of ASD. The participant completed the full task with minimal modification. The behavioural results showed that she understood the questions. The participant also did not report any undue distress during the experimental procedure.

Whilst this participant did not show a canonical N400 on S1, research has shown the N400 is present in both adults and children with a diagnosis of ASD (Wang et al., 2017). In neurotypical populations it is unlikely that all participants will show the effect. The study does however show this paradigm is feasible to do with more participants, and average results from a wider population might reveal more. Within a larger study, we would propose investigating the spatial and temporal properties of the N400 using a two-tailed cluster mass permutation test as in Chapter 7. Given an initial investigation should centre on the presence or absence of the Social N400, this should allow for differences in the N400 between Autistic and neurotypical populations. A further investigation could then ascertain the differences between matched groups (Autistic group compared to neurotypical control group matched on age, gender, language comprehension), something which cannot be achieved here through a single case investigation.

Participant I showed an interesting effect on sentence 5 in that the N400 appears inverted – the neural response to the PLAUS condition is more negative than to the IMPLAUS condition. Further analysis could show if this is consistent across other Autistic individuals. Inverted N400s have recently been shown in patients with schizophrenia with difficulties with language production – these patients showed a more negative response to primed words than non-primed (Kuperberg et al., 2018). However, it may also be the case that the strangeness of the experiment for the participant produced expectations which may have primed anticipation of IMPLAUS sentences rather than semantically congruent PLAUS sentences.

The study here is limited by the lack of independent assessment of Autism for the participant, instead relying on self-report of previous diagnosis. Further study could aim to standardise the assessment of Autism within the participant group, such as using the Autism Diagnostic Observation Schedule, Second Edition (ADOS-2). ADOS-2 has been shown to be highly accurate in identifying ASD, although it also shows a relatively high false-positive rate when used with individuals who have social communication difficulties through other dissociable conditions (Maddox et al., 2017). Combining this measure with greater understanding of previous healthcare assessment could better ensure validity and comparability of the experimental group.

In addition, the choice of a female participant here may limit the generalisability of results such that a further pilot with a male participant may be beneficial,

given the differences in both the N400 itself, and in presentation of Autism. The N400 has been shown to be a larger effect in female participants than male participants (Steffensen et al., 2008), which we might anticipate could also translate to the Social N400. Within ASD, there are significant sex/gender differences across elements which may be critical for this task such as executive function and Theory of Mind. Male Autistic participants perform more poorly on the Reading the Mind in the Eyes Task which may signal differences in using cues to determine agent actions, which our experiment in Chapter 5 might indicate would be a significant issue for interpreting the Joint Comprehension task as social (Holt et al., 2014). Female Autistic participants show greater difficulty with inhibitory control (Lemon et al., 2011), which might indicate differences in task performance or cue usage within Theory of Mind tasks. Overall, the limitation of this pilot to only one participant means that the impact of sex/gender differences in perception of the task has not yet been explored and may warrant further investigation prior to full-scale use of this Joint Comprehension task.

At first examination, the Joint Comprehension task appears to be feasible to use with participants on the Autism Spectrum. However, further investigation should be sensitive to sex/gender differences in the N400 and ASD; should investigate the N400 within Autistic and neurotypical populations; and should increase robustness of inclusion and matching criteria.

Chapter 9 Discussion

The body of work presented within this thesis aims to explore implicit Theory of Mind within a language task, asking: to what extent do individuals automatically process the comprehension of task partners during a Joint Comprehension task? Participants undertaking our modified within-subjects Joint Comprehension task showed clear usage of story context within early comprehension, modulating the N400 marker of incongruence. The same participants also showed a Social N400 when jointly reading a context-dependent sentence with a partner who has been deprived of context, even when the participant had full context for comprehension. The results of our investigation indicate several key findings:

- 1) The Social N400 demonstrates that individuals can represent the comprehension of joint readers within their own cognitive mechanisms for comprehension – we can represent others within the self in real time
- 2) The Social N400 does not appear to be explained by domain-general submentalising, and appears instead to be a social-specific effect
- 3) Contrary to a two-system model of Theory of Mind, representing others within the self (presumed to be the basis of implicit Theory of Mind) is not automatic or universal, but appears to be employed when it aids with task demands

- 4) Although not an automatic, universal system, the Social N400 correlates with previous studies of implicit Theory of Mind and suggests representing others within the self is used pre-adulthood.

Limitations of the current task

9.1.1 Sample size

A key criticism of the current body of work is the reliance on precedent for determination of sample size, with minimal participants recruited. This has in previous published literature been considered acceptable due to the robust nature of the effect being studied (the N400) (Bhavnani et al., 2021). However, there is growing concern about the comparability of results generated from small sample sizes compared to those from larger sample sizes (Baker et al., 2021), and for increased rigor when determining sample sizes and their impacts on effect sizes (Larson & Carbine, 2017; Šoškić et al., 2021).

When determining the necessary sample size for an experiment, we must also be cognizant that increasing sample size beyond necessary also increases the likelihood that an effect will be found (Baker et al., 2021). Therefore, future iterations of this experiment could look to use the results here as a pilot for determining sample size using power analysis. In our first experiment, testing the Social N400 using the new Joint Comprehension task in adult participants, we establish that the Social N400 has a mean of 1.34 and a standard deviation of 2.11.

Using these parameters, and a power of 0.8 (or 80% probability of finding the effect if there is an effect to be found), for a directional one-sample t-test, power analysis indicates a sample size of 17. For a bi-directional one-sample t-test as carried out in Chapter 6 where there was no strong hypothesis regarding presence or absence of the Social N400, power analysis indicates a sample size of 22 for an 80% probability of finding an effect on S5. The experiments outlined here use sample sizes smaller than indicated, and do not modulate sample size based on the directionality of analysis (directional analysis: N = 15, N = 14, N = 11, N = 16; bi-directional analysis: N = 15). Although large enough to detect a significant N400 on S1, it does raise the question both as to the size of the effects seen, and as to whether the Social N400 might be a weaker response requiring larger sample sizes for reliable detection.

The results outlined within this body of work therefore provide indication as to the direction of results, necessarily limited by the power due to the small sample sizes. Bayes analysis of null results on S5 indicate stronger support for accepting the null under the models that 1) the Social N400 requires mentalising, and 2) the Social N400 is not automatic; and weaker support under the model that break effects do not adequately explain the Social N400. An expansion of Chapter 4 could provide greater clarity of the potentially ambiguous results here, particularly given the visual similarities on S5 between Chapter 3 and Chapter 4.

9.1.2 Stimuli

Further refinement of the current Joint Comprehension task could seek to greater control the full stimuli set for differences between conditions. Target words in both PLAUS and IMPLAUS conditions were matched such that the same noun is used in each condition. The aim here was to limit the differences between PLAUS and IMPLAUS that could explain an N400 – such as the readability or recognisability of the critical words.

However, the same matching has not been applied to the full sentences that create the five-sentence stories. This introduces the possibility that there are differences within the sentence structure outside of the critical noun that might explain voltage differences between PLAUS and IMPLAUS. The difficulty of sentences prior to the critical noun might lead to differences in cognitive load between conditions. In addition, the stimuli are not controlled for valence, which might particularly lead to the potential for IMPLAUS stories to cause greater arousal (Delaney-Busch & Kuperberg, 2013).

Some comfort for the limited impact of these uncontrolled variables might come from the body of work as a whole – in that the presence or absence of an N400 on S5 can be modulated despite using identical stimuli. This suggests that uncontrolled differences between PLAUS and IMPLAUS stories may not be a primary indicator of the difference between PLAUS and IMPLAUS neural responses on S5. However, greater control of these variables can only lead to

greater confidence in results, given the potential of the task to prove a useful tool in exploring previously untapped cognitive mechanisms.

9.1.3 “Peeking”

As noted in the description of the Joint Comprehension task in Chapter 2, the task itself is reliant on the participant trusting that the confederate is not able to see the first four sentences of each story. It also requires the participant to understand that both parties are reading the final sentence of the story at the same time. In each experiment with a human confederate, we attempted to maximise this trust through visual cues, with the confederate raising their hands to cover their eyes and lowering their hands to uncover their eyes. However, there is no reason for the participant to trust that the confederate is not ‘peeking’ through their fingers.

To consider the impact of this on the results, we must consider how ‘peeking’ would adjust the confederate’s perspective of the task. If the confederate were to ‘peek’ for sentences 1 – 4, they would have the same story context as the participant on sentence 5. If participants were suspicious of this but still mentalising, we might therefore expect no N400 on S5 in all experiments – this could be considered equivalent to the experiment outlined in Chapter 4 in which the confederate is allowed to read sentences 1 – 4 with the participant but does not read sentence 5.

Outside of mentalising, it could be that the act of peeking vs not peeking modulates the arousal of the participant or the perception of the novelty of the task, which may in turn affect the N400 (Delaney-Busch & Kuperberg, 2013; Jacob & Huber, 2020). Given that the confederate's attention is the key task change on S5, inference about whether they have peeked might modulate the participant's response here. It's not clear the impact this would have on the N400 at S5. Current participants did not indicate they thought the confederate was peeking, but future iterations may wish to provide a more occlusive barrier to remove this perception. A variation in which peeking is included and made obvious to the participant might shed greater light on the potential impact at S5.

9.1.4 Reliance on language

The Joint Comprehension task in its current form was born out of investigations into context effects on language comprehension, and necessarily is a heavily language-related task, not a pure Theory of Mind task. Whilst not in itself an issue, given the strong intercorrelation shown between language and Theory of Mind (Gernsbacher & Yergeau, 2019; Happé & Frith, 2014; Happé, 1994), the level of language skill needed for success in the task could limit the use of the paradigm for further study.

Performance on Theory of Mind tasks has been shown to be modulated by language ability where there is a strong language component of the task

(Happé & Frith, 2014; Happé, 1994) highlighting the need to separate the two within task demands. Individuals who do not have the language skill necessary to follow the five-sentence story would not be able to be tested with the current paradigm. The potential for the future, therefore, is to investigate whether we can keep the structure of the current task but reduce the reliance on word-based stories. There is significant opportunity here, as the N400 is a marker of a mismatch between expectation and reality. Within language tasks, it is hypothesised to represent the extra cognitive resource required to discard predicted upcoming words in a sentence, and retrieve the actual word presented (Kuperberg et al., 2018; Mantegna et al., 2019). However, the effect is also shown across situations where similar mismatches occur – for example, in response to the misnaming of objects (Forgács et al., 2020; Vavatzanidis et al., 2018) and in response to mismatches between sounds and objects (Dudschig et al., 2018).

Given the potential for the Joint Comprehension task to help us better understand implicit Theory of Mind across populations, developing the task with a non-language analogue can only serve to increase the insight we can glean.

Developing a model of Theory of Mind

At the outset of this research, a prevalent model of Theory of Mind was the two-system model (Apperly & Butterfill, 2009). As discussed previously, this

model rests on two key assumptions: 1) that implicit Theory of Mind is automatic, and 2) that implicit and explicit Theory of Mind are underpinned by separate systems which follow temporally, where implicit comes first and explicit follows. Inconsistent support for these assumptions within behavioural data have previously been hypothesised to indicate that implicit Theory of Mind does not represent mentalising at all but is instead explained by domain-general submentalising (Heyes, 2014).

If the two-system model were to hold, we would anticipate that the Social N400 is always found in the presence of a social partner co-reading with the participants, regardless of task demands. However, this has not been shown to be the case: where the task does not prompt participants to consider the perspective of their co-reader, the Social N400 is absent (Chapter 6). However, the current results also do not support the suggestion that implicit Theory of Mind is explained entirely by submentalising. The Social N400 appears to be uniquely prompted by an agent with a mental state (i.e., comprehension capabilities), over and above domain-general properties such as attention signalling and event recording (Chapter 5).

The work here leads us to present an alternative model which may explain both the evidence presented within this thesis, and behavioural evidence within Theory of Mind to date. If we conceptualise Theory of Mind not as a rigid, dedicated system, but instead as an ability which is achieved through the dynamic recruitment of underlying capabilities (Siegal & Varley, 2002), our

hypothesis regarding implicit Theory of Mind changes. Could implicit Theory of Mind be considered a tool or mechanism which can be used when the task demands but competes with other mechanisms such that it is not recruited unless task demand is high enough. This would align with competition models of language processing (MacWhinney, 1992) but also would align with evidence that the brain seeks to address tasks in a resource-efficient manner (Skipper, 2015; Skipper & Lametti, 2021). Rather than a resource-expensive consistently available implicit Theory of Mind system, by weighting cues to task demands, unnecessary metabolic resource uses are avoided.

A model of Theory of Mind based on competition and dynamic recruitment of mechanisms helps us to interpret the pattern of results seen within this body of work, but also within the wider literature. Implicit Theory of Mind cannot be said to be automatic, based on the evidence from Chapter 6, and the inconsistent replicability of evidence such as looking patterns (Burnside et al., 2018; Kulke et al., 2018). However, where implicit Theory of Mind is said to be shown, the properties of it being fast, largely unconscious, and based on embodiment of the mental states of others within the self, does seem to hold consistently (Chapter 3) (Frith & Frith, 2008b). The term “implicit” therefore continues to suit, in as much as it is limited to “suggested but not communicated directly” (Cambridge Dictionary, n.d.) and not “complete and without any doubts”. Results from Chapter 5 of this thesis also lend weight to the assertion that this mechanism is truly social, in that it appears to require

an agent with a mental state to be prompted. As such, the Social N400 appears to be a candidate for a marker of true implicit Theory of Mind under the proposed model.

9.1.5 Threshold for implicit Theory of Mind

One of the most interesting quandaries that a competition model of Theory of Mind raises surrounds the threshold at which implicit Theory of Mind might be used, and the differences in this threshold across populations. For example, results from Chapter 5 indicate the Social N400 is reliant on the presence of an agent with a mental state, whereas Chapter 6 indicates that the Social N400 is not prompted by the mere presence of that agent where the task lacks any prompt to consider their mental state. This raises the question: is a task in which a co-actor is merely present a social task? Or indeed, at what level of interaction is the cue weight threshold met for implicit Theory of Mind to be used?

If we follow the logic from Bowler et al., (2005), a social task requires awareness of several elements, with the most relevant here being: 1) an agent, 2) acting towards a goal, 3) who has their own mental state that may be different to the participant. The Joint Comprehension task prompt “do you think the last sentence was plausible for your partner” indicates to the participant 1) the confederate is present, 2) the confederate is reading the sentence, 3) the confederate can or cannot understand the sentence, and that this may be

different to the participant's own experience. In taking out the question prompt, the task may fail at (3) as there is no need to have awareness of the agent's mental state. Although another person is present, at that point the task is no longer social, and the task may become more equivalent to that with the camera-agent in Chapter 5.

Modulating awareness of the confederate, their goals, and their potential mental state could be interesting for establishing the threshold at which a task becomes social, and thus implicit Theory of Mind might be recruited. For example, humans are wired to seek feedback within conversation, either verbal or non-verbal (Barr & Keysar, 2006). This feedback forms a critical element of the success of a conversation, and speakers often edit their own utterances based on feedback – aiding their listeners with the goal of understanding. Audience-design is hypothesised to be dependent on both implicit and explicit Theory of Mind, and therefore must involve mentalising (Hamilton & Lind, 2016; Horton & Gerrig, 2005). It could be interesting to explore whether feedback (e.g., laughter) from the confederate is enough to prompt awareness of the confederate, their goal, and their mental state, and to bring the task into the social realm, or whether there is a higher threshold of needing to *use* the co-reader's goal to be met for recruitment of social mechanisms.

In our trial outlined within Chapter 8, Participant I remarked that the confederate's task to simply sit and periodically read the screen seemed

incredibly boring. In such a framework, it could be easy to discount the confederate, and their goal, particularly in the absence of explicit questioning. Another option to investigate could be whether making the confederate's goal more explicit – i.e., giving them a concurrent task with a button press – might also increase the social nature of the task and the weighting applied to considering the confederate's perspective, particularly if both tasks hinge on the comprehension of sentence 5.

The results of Chapter 6 in fact contrast with evidence that the Social N400 is shown in the absence of explicit questioning about the confederate's perspective (Jouravlev et al., 2019). However, in the current study the confederate had no task beyond reading, whereas in the Jouravlev et al (2019) study the confederate had a button box and was asked to answer the questions the same as the participant. Evidence from joint action tasks demonstrate potential representation of others when their task is known about but not seen - indicating that it is the conceptualisation of the co-actor's goal which may be the key factor in use of implicit mentalising (Sebanz et al., 2005). This has perhaps in previous literature been misinterpreted as an effect of mere presence, but a co-actor with a goal could be a step beyond mere presence, and the goal may be the critical factor engaging social systems.

9.1.6 Implicit vs Explicit Theory of Mind

A necessary assumption from a competition model of Theory of Mind is that in order to successfully achieve Theory of Mind, multiple cognitive mechanisms must coordinate. Previous research has grouped mechanisms for Theory of Mind into implicit and explicit, with the latter assumed to rely on more reason-based mechanisms reliant on higher-order cognition (Apperly & Butterfill, 2009; Frith & Frith, 2008b). However, whilst outside use of these mechanisms might appear distinct, at least on some level they rely on common brain areas (Van Overwalle & Vandekerckhove, 2013) and common tools such as executive functions (Cane et al., 2017; Schneider, Lam, et al., 2012).

In previous research, behavioural measures of implicit Theory of Mind such as looking patterns have been disrupted by higher cognitive load (Schneider, Lam, et al., 2012). It would be interesting to explore whether the Social N400 is similarly affected by changes in cognitive load whilst completing the Joint Comprehension task, and further whether it is correlated with measures of executive functions. Previous research also suggests that, whilst implicit Theory of Mind may be recruited where there is motivation to do so, there could be a limit on available cognitive resources such that where participants are under a higher cognitive load, implicit Theory of Mind is abandoned in favour of prioritising resource for explicit Theory of Mind processes (Cane et al., 2017). Evidence from the Jouravlev et al (2019) studies already indicates that the Social N400 is in some cases disrupted by increased cognitive load

(Jouravlev et al., 2019). However, the Social N400 is displayed even when there is no prompting to consider the confederate's perspective. Could this be interpreted through the differing cognitive loads already involved within the tasks (5 sentences vs 2 sentences) meaning the threshold to use the Social N400 is higher in our variation of the task given potential resource constraints? We could therefore look to test whether there is a function of constraint, at which point implicit Theory of Mind is less likely to be used as the threshold becomes prohibitive, and whether this differs across populations.

It would also be interesting to understand the nature of the coordinating relationship between implicit and explicit Theory of Mind – does the embodiment of perspective feed into a rule-based approach, as rules are often developed based on experience? This could maintain a “first one, then the other” approach despite both being tools of the same system, and would indicate in some cases explicit Theory of Mind may be dependent on implicit Theory of Mind. Alternatively, is an answer to the task produced by both tools simultaneously, and then one discarded if they are in conflict, as suggested by previous research in children (Nadig & Sedivy, 2002)? This also raises further questions which may be outside of the scope of this task regarding how differing explicit mechanisms compete or coordinate with one another, as much of the literature to date has focussed on diverging answers generated implicit vs explicit mechanisms and not explicit vs explicit mechanisms.

9.1.7 Implicit Theory of Mind over the lifespan

As discussed earlier in the chapter, developing an analogous Joint Comprehension task outside of language opens up the ability to assess implicit Theory of Mind across populations. In particular, the ability to compare the Social N400 over the lifespan would be invaluable. The N400 itself is shown in infants as young as 6 months (Southgate & Verneti, 2014), leaving open the possibility of exploring Theory of Mind where current methods rely on looking patterns.

Indeed, a variation of the task has shown a Social N400 effect in infants as young as fourteen-months. Infants learned the label for an object, and then adults had a false belief about the label of the object induced. Although the label was congruous for the infants, they displayed a Social N400 as if the label was incongruous, in the same way it was for the adult (Forgács et al., 2020).

The ability to examine the Social N400 across the lifespan is particularly interesting within the context of a competition model of Theory of Mind, and raises the question: is age a moderator of cue weight? For example, looking patterns have been theorised to represent implicit Theory of Mind in infants without apparent motivation (Senju et al., 2010; Southgate et al., 2007; Southgate & Verneti, 2014), but may be reliant on task motivation in adults (Cane et al., 2017). Further, adolescents require less social interaction to build shared referents compared to adults (Kronmüller & Barr, 2007; Matthews et

al., 2010). Both findings appear to indicate that pre-adulthood, implicit Theory of Mind is used less sparingly than in adulthood. This is perhaps unsurprising if we consider the weight potentially applied to social cues within infancy, childhood, and adolescence. Infants are reliant on caregivers, and potentially more receptive to social cues as necessary for survival (Barry et al., 2015; Guellai & Streri, 2011). Adolescents are undergoing a huge amount of social change as they learn how they fit in to a social world without the direction of caregivers (Blakemore, 2018a, 2018b). Pre-adulthood, individuals have both a large capacity for, and a large need for, learning about the world around them, often facilitated through the people they interact with; language learning itself is hypothesised to be reliant on external social interaction which is internalised (Vygotsky, 1978, as cited in Goodman, 1990). Does this therefore change the weighting for recruitment of implicit Theory of Mind, or the threshold at which a task is considered social? When does this begin to change, and the weighting applied to implicit Theory of Mind reduce?

It could be interesting to run the Joint Comprehension task without the prompt to consider the confederate (as in Chapter 6) with adolescent participants. Given the potential for implicit Theory of Mind to be used less sparingly in adolescence, we might anticipate that the Social N400 is present without prompting in adolescent participants where it isn't in adults.

This hypothesis could also have some interesting implications for the groups deemed “adults” within many psychological studies. Adult participant

populations are often selected through university participant pools – these participant pools are often younger, but also often undergraduate students who are leaving home for the first time, meaning there could be a significant impact of social change. The relationship between age as a cue modulator, and the relationship between age and importance of social change, may need to be investigated separately and controlled for when aiming to consider each in isolation.

9.1.8 Theory of Mind in Autism Spectrum Disorder

The conceptualisation of a two-system model of Theory of Mind was, as discussed in Chapter 1, borne largely out of a model of impairment of Theory of Mind within ASD, and a potential double dissociation with neurotypical development. To step back from this model and instead consider a dynamic and flexible system of recruitment raises the potential for Theory of Mind to present, and be achieved, differently across individuals and across populations.

There is a wealth of evidence already that performance on many Theory of Mind tasks is related to language abilities, particularly vocabulary and grammar comprehension (Loukusa et al., 2014). As verbal mental age increases, performance on such tasks also increases (Happé & Frith, 2014). Research to date has shown that explicit Theory of Mind tasks are not predictive of a diagnosis of ASD, of empathy or emotional understanding, of

everyday social skills, or of peer relations (for review, see Gernsbacher & Yergeau, 2019). Despite this, social communication and interaction difficulties persist between those on the Autism Spectrum and neurotypical individuals. This has led to the suggestion that perhaps it is the faster-paced implicit Theory of Mind which underpins ‘success’ in general social interaction (Schaller & Rauh, 2017).

The Joint Comprehension task could play two roles here in testing whether implicit Theory of Mind is the root of dynamic social interaction in the everyday world, and whether the absence of such a mechanism is the cause of potential difficulties with this social interaction. Firstly, the Social N400, if a true measure of implicit Theory of Mind, as the current body of work appears to show, should be predictive of success in social interaction. One suggestion could be to determine whether the Social N400 is predictive of scores in other social measures such as measures of empathy. Secondly, absence of the Social N400 should be predictive of difficulties in social interaction or reasoning. Within the confines of the current task – could presence or absence of the Social N400 predict performance on Q1: Do you think the last sentence was plausible for your partner? Or further, is the absence of a Social N400 predictive of a diagnosis of ASD?

Reconceptualising the model of Theory of Mind opens several other options here, outside of a model of ‘impairment’. Given individuals on the Autism Spectrum often pass Theory of Mind tasks (Happé & Frith, 2014), the concept

of a difference between those with ASD and neurotypical individuals might not be that of 'whole' and 'impaired' but instead a matter of differences in recruitment of cognitive mechanisms. Relevant here is also the concept of the "double empathy problem" between social actors within differing outlooks (Milton, 2012) – there is an issue from both neurotypical and Autistic individuals in conceptualising the mental state of the other, but it is only the Autistic point of view which is pathologized. The pathway from 'impairment' to 'it works, but differently' is not unique to Theory of Mind, and stems from an approach of 'normocentrism' (Mottron et al., 2008) in which researchers attempt to map autistic cognitive processing onto models of neurotypical cognitive processing, with the propensity to then interpret differences as deficits. Outside of a model of impairment of implicit Theory of Mind, might there instead be differences in the way Theory of Mind is constructed, and the way implicit Theory of Mind might be recruited.

One hypothesis could be that Autistic individuals might differ in their use of cue weights to recruit cognitive mechanisms for Theory of Mind. It could be that implicit Theory of Mind is used by Autistic individuals, particularly in a task where perspective-taking is so explicitly prompted, and so a Social N400 could be expected, but that the modulation of this effect based on task prompting shows a different profile to neurotypical participants. If this were the case, and implicit Theory of Mind were also shown to support everyday social interaction, incremental support could be trialled to increase the weight of the cue for

implicit Theory of Mind usage. Task support has been shown to help with task performance in other areas such as memory (Bowler et al., 2015). Within social interaction with neurotypical individuals, often the expectation to consider others' perspectives is not explicitly communicated (there is no "Question 1") and additional support to recognise such a mechanism would aid with understanding goals and intentions could cue an entirely different cognitive approach.

Examples of everyday unsupported cues to consider perspectives within language include: metaphors (Bambini et al., 2016; Leslie & Frith, 1988); sense of humour (Silva et al., 2017); white lies (Moreno et al., 2016). These are complex concepts which can all be used flexibly and show key differences in interpretation between neurotypical and Autistic individuals. These structures have been shown to lead to N400 responses when expectations are violated in neurotypical participants. Future studies could examine whether the N400 in response to metaphors can be modulated by context, and whether this is then impacted by the context that a conversation partner has, and whether this cue use is modulated differently, or interacts differently with other cues, in Autistic individuals. A greater understanding of the ways in which multiple mechanisms are recruited to achieve a task, and the interaction between these mechanisms can only lead to greater understanding of the complex dynamic process of social cognition.

Combined with study into the Social N400 across the lifespan, an adjusted paradigm might provide a useful tool for early intervention within ASD. Greater understanding of differences in patterns of tool use, and the support for prompting tool use, could provide evidence for supportive practices within early life that may aid with development. For example, if we can establish differences in minimal thresholds to cue implicit Theory of Mind, one could argue that use of the higher support could aid across situations such as classroom-based learning, particularly if it could boost language learning for children who might be disadvantaged by a neurotypical framework (McIntyre et al., 2018). Evidence from a paradigm such as that by Forgács et al., (2020) could also inform cue support within early infancy by caregivers. Understanding how cues are weighted and how they interact can inform our support for a wider population, beyond the hints and assumptions the social world is loaded with for now.

Conclusion

The body of work outlined here aims to use a Joint Comprehension task to explore implicit Theory of Mind, and the way in which individuals consider the perspectives of others. Individuals appear to model the comprehension of task partners within their own comprehension mechanisms, in real time, demonstrated with a Social N400 effect. This effect appears to be contingent on the need to consider the perspective of an agent (a co-reader) with a different mental state to the participants' own but does not appear without

prompting to consider the co-reader. The Social N400 effect is described within the context of a competition model of Theory of Mind, suggesting that implicit Theory of Mind is a cognitive tool which can be used when it would aid with task demands. This could have an impact on our understanding of Theory of Mind across the lifespan in both neurotypical and Autistic individuals, and could lead to future tools for support with social interaction.

Appendix

PLAUS	1	The man	bought the	banana.	2	The banana	was	ripe.	3	The banana	had been	imported.	4	The man	peeled the	banana.	5	The man	ate the	banana.
PLAUS	1	The woman	boiled the	cabbage.	2	The cabbage	was finely	chopped.	3	The cabbage	was	fresh.	4	The woman	poured gravy on the	cabbage.	5	The woman	ate the	cabbage.
PLAUS	1	The boy	rode the	bike.	2	The bike	had a flat	tire.	3	The boy	stopped the	bike.	4	The boy	got off the	bike.	5	The boy	pushed the	bike.
PLAUS	1	The woman	cooked the	steak.	2	The steak	was from the	butchers.	3	The woman	added sauce to the	steak.	4	The woman	fried the	steak.	5	The woman	tasted the	steak.
PLAUS	1	The girl	picked the	flower.	2	The girl	put the flower in her	hair.	3	The flower	smelled	nice.	4	The girl	liked the	flower.	5	The girl	admired the	flower.
PLAUS	1	The boy	decorated the	frame.	2	The frame	was to be a	present.	3	The frame	looked	dirty.	4	The boy	cleaned the	frame.	5	The boy	wrapped up the	frame.
PLAUS	1	The boy	flipped the	pancake.	2	The pancake	was round and	sweet.	3	The boy	spread syrup on the	pancake.	4	The boy	tasted the	pancake.	5	The boy	ate the	pancake.
PLAUS	1	The girl	peeled the	orange.	2	The orange	was very	juicy.	3	The orange	dripped on the girl's	clothes.	4	The girl	tasted the	orange.	5	The girl	ate the	orange.
PLAUS	1	The boy	stared at the	moon.	2	The moon	was	big.	3	The moon	was	yellow.	4	The boy	admired the	moon.	5	The boy	pointed at the	moon.
PLAUS	1	The woman	viewed the	flat.	2	The woman	liked the	flat.	3	The woman	asked the price of the	flat.	4	The woman	rented the	flat.	5	The woman	moved into the	flat.
PLAUS	1	The woman	watered the	plant.	2	The plant	grew	strong.	3	The plant	needed water and	sunlight.	4	The woman	pruned the	plant.	5	The woman	re-potted the	plant.
PLAUS	1	The boy	bought the	notebook .	2	The notebook	had lined	pages.	3	The notebook	was	empty.	4	The boy	wrote in the	notebook .	5	The boy	filled the	notebook .
PLAUS	1	The woman	looked up the	number.	2	The number	had seven	digits.	3	The number	was for a	restaurant.	4	The woman	dialled the	number.	5	The woman	remembered the	number.
PLAUS	1	The woman	bought the	berries.	2	The berries	were	blue.	3	The woman	washed the	berries.	4	The woman	cooked with the	berries.	5	The woman	liked the taste of the	berries.
PLAUS	1	The man	decorated the	bedroom .	2	The man	painted the bedroom	blue.	3	A bed	was moved in to the	bedroom.	4	The man	was proud of the	bedroom .	5	The man	slept in the	bedroom .
PLAUS	1	The boy	wanted the	balloon.	2	The boy	went to find the	balloon	3	The boy	purchased the	balloon.	4	The boy	liked the	balloon.	5	The boy	played with the	balloon.

PLAUS	1	The woman	wore the	hat.	2	The hat	was	yellow.	3	The hat	fell on the	floor.	4	The woman	dusted off the	hat.	5	The woman	put on the	hat.
PLAUS	1	The lady	ran out of the	jam.	2	The lady	bought more of the	jam.	3	The jam	was the lady's	favourite.	4	The lady	ate toast with the	jam.	5	The lady	ate lots of the	jam.
PLAUS	1	The woman	purchased the	house.	2	The woman	moved into the	house.	3	The house	was	spacious.	4	The woman	enjoyed the	house.	5	The woman	decorated the	house.
PLAUS	1	The man	played the	piano.	2	The piano	was old and	broken.	3	The man	repaired the	piano.	4	The man	practiced the	piano.	5	The man	was talented at the	piano.
PLAUS	1	The woman	bought the	cooker.	2	The cooker	had lots of	shelves.	3	The cooker	was	expensive.	4	The woman	cleaned the	cooker.	5	The woman	baked in the	cooker.
PLAUS	1	The girl	drank the	milk.	2	The milk	tasted	strange.	3	The milk	was out of	date.	4	The girl	did not drink the	milk.	5	The girl	poured away the	milk.
PLAUS	1	The woman	opened the	wine.	2	The woman	poured the wine into a	glass.	3	The woman	smelled the	wine.	4	The woman	tasted the	wine.	5	The woman	liked the	wine.
PLAUS	1	The woman	needed the	sugar.	2	The sugar	was for a	cake.	3	The woman	couldn't bake without the	sugar.	4	The woman	searched for the	sugar.	5	The woman	couldn't find the	sugar.
PLAUS	1	The girl	played on the	swing.	2	The girl	enjoyed playing on the	swing.	3	The swing	broke under the	girl.	4	The girl	felt guilty about the	swing.	5	The girl	fixed the	swing.
PLAUS	1	The man	looked at the	tree.	2	The tree	was	overgrown.	3	The man	decided to remove the	tree.	4	The man	cut down the	tree.	5	The man	chopped up the	tree.
PLAUS	1	The man	searched for the	spoon.	2	The man	wanted a	spoon.	3	The man	tried to find the	spoon.	4	The man	found the	spoon.	5	The man	used the	spoon.
PLAUS	1	The woman	bought the	potato.	2	The woman	wanted potato for	dinner.	3	The woman	cooked the	potato.	4	The woman	seasoned the	potato.	5	The woman	ate the	potato.
PLAUS	1	The man	put on the	trousers.	2	The trousers	were too	small.	3	The trousers	were	scruffy.	4	The man	took off the	trousers.	5	The man	donated the	trousers.
PLAUS	1	The woman	bought the	scissors.	2	The scissors	were	expensive.	3	The scissors	were very	sharp.	4	The woman	used the	scissors.	5	The woman	was careful with the	scissors.
PLAUS	1	The girl	made the	pizza.	2	The pizza	had many	toppings.	3	The girl	wanted the	pizza.	4	The girl	tried some of the	pizza.	5	The girl	liked the	pizza.
PLAUS	1	The boy	ate the	pear.	2	The pear	was	green.	3	The pear	was sweet and	juicy.	4	The boy	polished the	pear.	5	The boy	enjoyed eating the	pear.
PLAUS	1	The child	kicked the	ball.	2	The ball	was very	bouncy.	3	The child	played with the	ball.	4	The child	made friends using the	ball.	5	The child	liked playing with the	ball.
PLAUS	1	The boy	bought a	cactus.	2	The cactus	was green and	spikey.	3	The cactus	was very	small.	4	The boy	watered the	cactus.	5	The boy	looked after the	cactus.

PLAUS	1	The man	cooked the	egg.	2	The egg	was fresh from the	farm.	3	The egg	was very	large.	4	The man	fried the	egg.	5	The man	had bacon with the	egg.
PLAUS	1	The man	owned the	tractor.	2	The tractor	was a	green.	3	The tractor	attracted lots of	visitors.	4	People	took rides on the	tractor.	5	Everyon e	enjoyed rides on the	tractor.
PLAUS	1	The man	ate the	grapes.	2	The grapes	were	seedless.	3	The grapes	were	imported.	4	The man	washed the	grapes.	5	The man	enjoyed eating the	grapes.
PLAUS	1	The boy	decided to eat the	sausage.	2	The sausage	was from the	butchers.	3	The sausage	was made with	pork.	4	The boy	cooked the	sausage.	5	The boy	ate the	sausage.
PLAUS	1	The friend	cooked the	tomato.	2	The tomato	needed to be	eaten.	3	The tomato	was still	ripe.	4	The friend	chopped the	tomato.	5	The friend	made sauce with the	tomato.
PLAUS	1	The girl	looked up at the	cloud.	2	The cloud	looked grey and	stormy.	3	The cloud	was going to	rain.	4	The girl	looked at the	cloud.	5	It rained	heavily from the	cloud.
PLAUS	1	The museum	displayed the	bones.	2	The bones	were as big as a	person.	3	The bones	were very	old.	4	The students	visited the	bones.	5	The student s	studied the	bones.
PLAUS	1	The grandma	baked the	apples.	2	The apples	were wrapped in	pastry.	3	The apples	were	sugary.	4	The grandma	prepared the	apples.	5	The grandm a	liked baking the	apples.
PLAUS	1	The Queen	wanted to eat the	scone.	2	The Queen	had saved the	scone.	3	The Queen	enjoyed the	scone.	4	The Queen	ate the	scone.	5	The Queen	put jam on the	scone.
PLAUS	1	The boy	polished the	guitar.	2	The guitar	was brand	new.	3	The boy	adored the	guitar.	4	The boy	played rock on the	guitar.	5	The boy	practised the	guitar.
PLAUS	1	The woman	wore the	bracelet.	2	The bracelet	was	golden.	3	The bracelet	was a	present.	4	The woman	showed her friends the	bracelet.	5	The woman	liked the	bracelet.
PLAUS	1	The musician	played the	flute.	2	The musician	enjoyed the	flute.	3	Playing the	flute calmed the	musician.	4	The musician	performed with the	flute.	5	The musician	played perfectly on the	flute.
PLAUS	1	The man	read the	newspap er.	2	The newspaper	was	depressin g.	3	The newspap er	covered important	events.	4	The man	liked reading the	newspap er.	5	The man	folded the	newspap er.
PLAUS	1	The man	filled the	kettle.	2	The man	turned on the	kettle.	3	The man	boiled the	kettle.	4	The man	poured water from the	kettle.	5	The man	refilled the	kettle.
PLAUS	1	The boy	disliked the	sprouts.	2	The sprouts	tasted	bad.	3	The sprouts	didn't look	tasty.	4	The boy	threw away the	sprouts.	5	The boy	did not eat the	sprouts.
PLAUS	1	The builder	built the	walls.	2	The walls	were very	sturdy.	3	The walls	were made from	bricks.	4	The builder	was pleased with the	walls.	5	The builder	perfected the	walls.
PLAUS	1	The cleaner	used the	broom.	2	The cleaner	swept with the	broom.	3	The broom	was	large.	4	The cleaner	broke the	broom.	5	The cleaner	replaced the	broom.
PLAUS	1	The boy	tried to crack the	coconuts.	2	The coconuts	were	large.	3	The coconuts	were hard to	break.	4	The boy	split the	coconuts.	5	The boy	drank from the	coconuts.

IMPLA US	1	The book	was written by the	banana.	2	The banana	became	famous.	3	The banana	earned a lot of	money.	4	People	recognised the	banana.	5	A reporter	interviewed the	banana.
IMPLA US	1	The man	spoke to the	cabbage.	2	The cabbage	was	hungry.	3	The cabbage	had been working	hard.	4	The man	made food for the	cabbage.	5	The man	fed the	cabbage.
IMPLA US	1	The girl	argued with the	bike.	2	The girl	made the bike	angry.	3	The bike	was scared of the	girl.	4	The girl	threatened the	bike.	5	The girl	hurt the	bike.
IMPLA US	1	The man	humiliated the	steak.	2	The steak	felt	ashamed	3	The steak	confronted the	man.	4	The man	apologised to the	steak.	5	The man	became friends with the	steak.
IMPLA US	1	The dancer	taught the	flower.	2	The dancer	thought the flower was	talented.	3	The flower	did not enjoy	dancing.	4	The dancer	was frustrated with the	flower.	5	The dancer	struggled to teach the	flower.
IMPLA US	1	The boy	read to the	frame.	2	The boy	let the frame pick the	story.	3	The frame	liked picture	books.	4	The boy	showed the pictures to the	frame.	5	The boy	enjoyed reading to the	frame.
IMPLA US	1	The boy	visited the	pancake.	2	The pancake	was alone at	home.	3	The pancake	had tidied the	house.	4	The boy	spoke to the	pancake.	5	The boy	was kind to the	pancake.
IMPLA US	1	The girl	tutored the	orange.	2	The orange	had trouble with	maths.	3	The orange	was confused by	numbers.	4	The girl	was patient with the	orange.	5	The girl	helped the	orange.
IMPLA US	1	The man	drove with the	moon.	2	The moon	had borrowed a	car.	3	The moon	was a bad	driver.	4	The car	was crashed by the	moon.	5	The man	sued the	moon.
IMPLA US	1	The woman	instructed the	flat.	2	The flat	was learning to	paint.	3	The flat	Painted a	masterpiece	4	The masterpiece	was admired by the	flat.	5	The woman	congratulated the	flat.
IMPLA US	1	The woman	dated the	plant.	2	The plant	loved the	woman.	3	The plant	proposed to the	woman.	4	The woman	rejected the	plant.	5	The woman	upset the	plant.
IMPLA US	1	The boy	invited the	notebook	2	The notebook	was very	excited.	3	The notebook	liked the	boy.	4	The boy	was polite to the	notebook	5	The boy	welcomed the	notebook
IMPLA US	1	The woman	danced with the	number.	2	The number	liked to	tango.	3	The number	could dance	well.	4	The woman	spun about with the	number.	5	The woman	tangoed with the	number.
IMPLA US	1	The girl	spoke to the	berries.	2	The berries	were	crying.	3	The berries	wiped their	tears.	4	The girl	cuddled the	berries.	5	The girl	cheered up the	berries.
IMPLA US	1	The woman	pitied the	bedroom	2	The bedroom	was	crying.	3	The bedroom	felt	lonely.	4	The woman	listened to the	bedroom	5	The woman	comforted the	bedroom
IMPLA US	1	The woman	spoke to the	balloon.	2	The balloon	got a lift to	school.	3	The balloon	didn't like	school.	4	The woman	forced the	balloon.	5	The woman	upset the	balloon.
IMPLA US	1	The man	raced the	hat.	2	The hat	was running	fast.	3	The man	was slower than the	hat.	4	The man	caught up with the	hat.	5	The man	lost to the	hat.
IMPLA US	1	The lady	excited the	jam.	2	The lady	took the jam to the	fair.	3	The jam	went on a	ride.	4	The lady	smiled at the	jam.	5	The lady	treated the	jam.

IMPLA US	1	The woman	spoke to the	house.	2	The house	felt	lonely.	3	The house	wanted a	hug.	4	The woman	hugged the	house.	5	The woman	comforted the	house.
IMPLA US	1	The man	cooked for the	piano.	2	The piano	felt	hungry.	3	The man	cooked the piano	pasta.	4	The man	served the	piano.	5	The woman The meal	filled up the	piano.
IMPLA US	1	The man	argued with the	cooker.	2	The cooker	looked	angry.	3	The cooker	attacked the	man.	4	The man	was insulted by the	cooker.	5	The man	was bullied by the	cooker.
IMPLA US	1	The girl	texted the	milk.	2	The milk	read the	text.	3	The milk	thought of an	answer.	4	The girl	heard from the	milk.	5	The girl	did not answer the	milk.
IMPLA US	1	The woman	chatted to the	wine.	2	The woman	and the wine were	dating.	3	The woman	listened to the	wine.	4	The woman	heard jokes by the	wine.	5	The woman	liked dating the	wine.
IMPLA US	1	The woman	read to the	sugar.	2	The sugar	liked	stories.	3	The sugar	preferred horror	stories.	4	The woman	always read to the	sugar.	5	The woman	loved reading to the	sugar.
IMPLA US	1	The girl	watched films with the	swing.	2	The swing	liked going out with	friends.	3	The swing	bought some	popcorn.	4	The girl	shared popcorn with the	swing.	5	The girl	liked sharing with the	swing.
IMPLA US	1	The woman	argued with the	tree.	2	The tree	felt very	upset.	3	The tree	hated the	woman.	4	The woman	felt sorry for the	tree.	5	The woman	consolated the	tree.
IMPLA US	1	The instruct or	taught the	spoon.	2	The spoon	liked	gymnastics.	3	The spoon	showed off to the	boy.	4	The boy	watched the	spoon.	5	The boy	applauded the	spoon.
IMPLA US	1	The woman	hurt the	potato.	2	The woman	teased the	potato.	3	The potato	felt	vulnerable.	4	The woman	upset the	potato.	5	The woman	apologised to the	potato.
IMPLA US	1	The man	chatted with the	trousers.	2	The trousers	felt	unwell.	3	The trousers	wanted to feel	better.	4	The man	cared for the	trousers.	5	The man	cheered up the	trousers.
IMPLA US	1	The woman	phoned the	scissors.	2	The scissors	loved	chatting.	3	The scissors	were very	talkative.	4	The woman	gossiped with the	scissors.	5	The woman	loved phoning the	scissors.
IMPLA US	1	The girl	had to feed the	pizza.	2	The pizza	hadn't eaten all	day.	3	The pizza	loved the girl's	cooking.	4	The girl	made food for the	pizza.	5	The girl	liked to cook for the	pizza.
IMPLA US	1	The teacher	tutored the	pear.	2	The pear	made lots of	notes.	3	The pear	was a good	student.	4	The teacher	gave help to the	pear.	5	The teacher	was proud of the	pear.
IMPLA US	1	The child	was kicked by the	ball.	2	The ball	had a nasty	temper.	3	The ball	was in a bad	mood.	4	The child's	money was stolen by the	ball.	5	The child	was bullied by the	ball.
IMPLA US	1	The boy	was friends with the	cactus.	2	The boy	saw a film with the	cactus.	3	The cactus	liked action	movies.	4	The boy	saw a comedy with the	cactus.	5	The film	was disliked by the	cactus.
IMPLA US	1	The teacher	spoke to the	egg.	2	The egg	had not done his	homework.	3	The egg	felt	stressed.	4	The work	overloaded the	egg.	5	The teacher	helped the	egg.
IMPLA US	1	The DJ	partied with the	tractor.	2	The tractor	was a good	dancer.	3	The tractor	met up with	friends.	4	The friends	cheered the	tractor.	5	The teacher Everyone	danced with the	tractor.

IMPLA US	1	The man	was offended by the	grapes.	2	The grapes	had used bad	language.	3	The grapes	wanted to cause	offence.	4	The man	told off the	grapes.	5	The man	punished the	grapes.
IMPLA US	1	The boy	phoned the	sausage.	2	The sausage	had a long	phone call.	3	The sausage	was speaking	loudly.	4	People	were annoyed with the	sausage.	5	A man	hushed the	sausage.
IMPLA US	1	The friend	argued with the	tomato.	2	The tomato	was having a bad	day.	3	The tomato	reacted	badly.	4	The friend	teased the	tomato.	5	The friend	apologised to the	tomato.
IMPLA US	1	The friend	had tea with the	cloud.	2	The cloud	chatted with the	friend.	3	The cloud	liked	biscuits.	4	The friend	made tea for the	cloud.	5	The friend	gave biscuits to the	cloud.
IMPLA US	1	The policeman	was annoyed by the	bones.	2	The bones	spray painted the	wall.	3	The bones	were	punished.	4	The policeman	was ignored by the	bones.	5	The policeman	arrested the	bones.
IMPLA US	1	The orchestra	auditioned the	apples.	2	The apples	were extremely	nervous.	3	The apples	played the	violin.	4	The orchestra	hired the	apples.	5	The orchestra	directed the	apples.
IMPLA US	1	The Queen	listened to the	scone.	2	The scone	was talking about	taxes.	3	The scone	was very	opinionated .	4	The Queen	did not trust the	scone.	5	The Queen	imprisoned the	scone.
IMPLA US	1	The man	cooked for the	guitar.	2	The guitar	felt	hungry.	3	The man	cooked the guitar	pasta.	4	The man	served the	guitar.	5	The meal	filled up the	guitar.
IMPLA US	1	The band	hired the	bracelet.	2	The bracelet	was a	pianist.	3	The bracelet	was good at the	piano.	4	The band	performed with the	bracelet.	5	The band	rocked out with the	bracelet.
IMPLA US	1	The lady	married the	flute.	2	The flute	wore a	suit.	3	The flute	prepared some	vows.	4	The lady	argued with the	flute.	5	The lady	divorced the	flute.
IMPLA US	1	The man	danced with the	newspaper.	2	The newspaper	was	talented.	3	The man	admired the	newspaper.	4	The man	waltzed with the	newspaper.	5	The man	proposed to the	newspaper.
IMPLA US	1	The man	confided in the	kettle.	2	The kettle	was a good	listener.	3	The man	found the kettle	understanding.	4	The man	told a secret to the	kettle.	5	The man	trusted the	kettle.
IMPLA US	1	The boy	skipped class with the	sprouts.	2	The sprouts	were	naughty.	3	The sprouts	loved causing	trouble.	4	The boy	was misled by the	sprouts.	5	The boy	was grounded with the	sprouts.
IMPLA US	1	The man	harassed the	walls.	2	The man	was taunted by the	walls.	3	The walls	were	rude.	4	The man	was insulted by the	walls.	5	The man	made enemies with the	walls.
IMPLA US	1	The cleaner	comforted the	broom.	2	The broom	was very	depressed.	3	The broom	had gone through a	breakup.	4	The girl	consoled the	broom.	5	The girl	was helping the	broom.
IMPLA US	1	The boy	was scared of the	coconuts.	2	The coconuts	were chasing the	boy.	3	The coconuts	had	teeth.	4	The boy	hid from the	coconuts.	5	The boy	escaped from the	coconuts.

Bibliography

- American Psychiatric Association. *Diagnostic and statistical manual of mental disorders*. 5th ed. Arlington, VA: American Psychiatric Association; 2013.
- Apperly, I. A., & Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review*, 116(4), 953–970. <https://doi.org/10.1037/a0016923>
- Baddeley, A. (2003). Working memory: Looking back and looking forward. *Nature Reviews. Neuroscience*, 4(10), 829–839. <https://doi.org/10.1038/nrn1201>
- Baker, D. H., Vilidaite, G., Lygo, F. A., Smith, A. K., Flack, T. R., Gouws, A. D., & Andrews, T. J. (2021). Power contours: Optimising sample size and precision in experimental psychology and human neuroscience. *Psychological Methods*, 26(3), 295–314. <https://doi.org/10.1037/met0000337>
- Bambini, V., Bertini, C., Schaeken, W., Stella, A., & Di Russo, F. (2016). Disentangling Metaphor from Context: An ERP Study. *Frontiers in Psychology*, 7, 559. <https://doi.org/10.3389/fpsyg.2016.00559>
- Baron-Cohen, S. (2000). Theory of mind and autism: A review. In *International Review of Research in Mental Retardation* (Vol. 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., & Frith, U. (1985). Does the autistic child have a “theory of mind”? *Cognition*, 21(1), 37–46.
[https://doi.org/10.1016/0010-0277\(85\)90022-8](https://doi.org/10.1016/0010-0277(85)90022-8)
- Barr, D. J., & Keysar, B. (2006). Chapter 23—Perspective Taking and the Coordination of Meaning in Language Use. In M. J. Traxler & M. A. Gernsbacher (Eds.), *Handbook of Psycholinguistics (Second Edition)* (pp. 901–938). Academic Press. <https://doi.org/10.1016/B978-012369374-7/50024-9>
- Barry, R. A., Graf Estes, K., & Rivera, S. M. (2015). Domain general learning: Infants use social and non-social cues when learning object statistics. *Frontiers in Psychology*, 6, 551.
<https://doi.org/10.3389/fpsyg.2015.00551>
- Bentin, S., McCarthy, G., & Wood, C. C. (1985). Event-related potentials, lexical decision and semantic priming. *Electroencephalography and Clinical Neurophysiology*, 60(4), 343–355.
[https://doi.org/10.1016/0013-4694\(85\)90008-2](https://doi.org/10.1016/0013-4694(85)90008-2)
- Bhavnani, S., Estrin, G. L., Haartsen, R., Jensen, S. K. G., Gliga, T., Patel, V., & Johnson, M. H. (2021). EEG signatures of cognitive and social development of preschool children—a systematic review. *PLOS ONE*, 16(2), e0247223. <https://doi.org/10.1371/journal.pone.0247223>
- Blair, R. C., & Karniski, W. (1993). An alternative method for significance testing of waveform difference potentials. *Psychophysiology*, 30(5), 518–524.

- Blakemore, S.-J. (2018a). *Inventing Ourselves: The Secret Life of the Teenage Brain*. Public Affairs Books.
- Blakemore, S.-J. (2018b). Avoiding Social Risk in Adolescence. *Current Directions in Psychological Science*, 27(2), 116–122. <https://doi.org/10.1177/0963721417738144>
- Blakemore, S.-J., & Choudhury, S. (2006). Development of the adolescent brain: Implications for executive function and social cognition. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 47(3–4), 296–312. <https://doi.org/10.1111/j.1469-7610.2006.01611.x>
- Bowler, D. M., Briskman, J., Gurvidi, N., & Fornells-Ambrojo, M. (2005). Understanding the Mind or Predicting Signal-Dependent Action? Performance of Children With and Without Autism on Analogues of the False-Belief Task. *Journal of Cognition and Development*, 6(2), 259–283. https://doi.org/10.1207/s15327647jcd0602_5
- Bowler, D. M., Gaigg, S. B., & Gardiner, J. M. (2015). Brief Report: The Role of Task Support in the Spatial and Temporal Source Memory of Adults with Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*, 45(8), 2613–2617. <https://doi.org/10.1007/s10803-015-2378-9>
- Brennan, S. E., Galati, A., & Kuhlen, A. K. (2010). Chapter 8 - Two Minds, One Dialog: Coordinating Speaking and Understanding. In B. H. Ross (Ed.), *Psychology of Learning and Motivation* (Vol. 53, pp. 301–344). Academic Press. [https://doi.org/10.1016/S0079-7421\(10\)53008-1](https://doi.org/10.1016/S0079-7421(10)53008-1)

- Brown-Schmidt, S. (2009). Partner-specific interpretation of maintained referential precedents during interactive dialog. *Journal of Memory and Language*, 61(2), 171–190. <https://doi.org/10.1016/j.jml.2009.04.003>
- Bullmore, E. T., Suckling, J., Overmeyer, S., Rabe-Hesketh, S., Taylor, E., & Brammer, M. J. (1999). Global, voxel, and cluster tests, by theory and permutation, for a difference between two groups of structural MR images of the brain. *IEEE Transactions on Medical Imaging*, 18(1), 32–42. <https://doi.org/10.1109/42.750253>
- Burnett, S., Bird, G., Moll, J., Frith, C., & Blakemore, S.-J. (2008). Development during Adolescence of the Neural Processing of Social Emotion. *Journal of Cognitive Neuroscience*, 21(9), 1736–1750. <https://doi.org/10.1162/jocn.2009.21121>
- Burnside, K., Ruel, A., Azar, N., & Poulin-Dubois, D. (2018). Implicit false belief across the lifespan: Non-replication of an anticipatory looking task. *Cognitive Development*, 46, 4–11. <https://doi.org/10.1016/j.cogdev.2017.08.006>
- Butterfill, S. A., & Apperly, I. A. (2013). How to Construct a Minimal Theory of Mind. *Mind & Language*, 28(5), 606–637. <https://doi.org/10.1111/mila.12036>
- Cane, J. E., 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(4), 591–610. <https://doi.org/10.1037/xlm0000345>

- Clark, H. H., & Krych, M. A. (2004). Speaking while monitoring addressees for understanding. *Journal of Memory and Language*, 50(1), 62–81.
<https://doi.org/10.1016/j.jml.2003.08.004>
- Colvin, M., & Warren, T. (2020). Language comprehension models at different levels of explanation. *Language and Linguistics Compass*, 14(1), e12362. <https://doi.org/10.1111/lnc3.12362>
- Conway, J. R., Lee, D., Ojaghi, M., Catmur, C., & Bird, G. (2017). Submentalizing or mentalizing in a Level 1 perspective-taking task: A cloak and goggles test. *Journal of Experimental Psychology. Human Perception and Performance*, 43(3), 454–465.
<https://doi.org/10.1037/xhp0000319>
- Craft, J. L., & Simon, J. R. (1970). Processing symbolic information from a visual display: Interference from an irrelevant directional cue. *Journal of Experimental Psychology*, 83(3), 415–420.
- Cummings, A., Ceponiene, R., Dick, F., Saygin, A. P., & Townsend, J. (2008). A developmental ERP study of verbal and non-verbal semantic processing. *Brain Research*, 1208, 137–149.
<https://doi.org/10.1016/j.brainres.2008.02.015>
- De Jaegher, H., Di Paolo, E., & Gallagher, S. (2010). Can social interaction constitute social cognition? *Trends in Cognitive Sciences*, 14(10), 441–447. <https://doi.org/10.1016/j.tics.2010.06.009>
- Delaney-Busch, N., & Kuperberg, G. (2013). Friendly drug-dealers and terrifying puppies: Affective primacy can attenuate the N400 effect in

- emotional discourse contexts. *Cognitive, Affective, & Behavioral Neuroscience*, 13(3), 473–490. <https://doi.org/10.3758/s13415-013-0159-5>
- Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, 56(1), 5–18. <https://doi.org/10.1037/0022-3514.56.1.5>
- Diamond, A. (2013). Executive Functions. *Annual Review of Psychology*, 64, 135–168. <https://doi.org/10.1146/annurev-psych-113011-143750>
- Dudschig, C., Mackenzie, I. G., Leuthold, H., & Kaup, B. (2018). Environmental sound priming: Does negation modify N400 cross-modal priming effects? *Psychonomic Bulletin & Review*, 25(4), 1441–1448. <https://doi.org/10.3758/s13423-017-1388-3>
- Dumontheil, I., Apperly, I. A., & Blakemore, S.-J. (2010). Online usage of theory of mind continues to develop in late adolescence. *Developmental Science*, 13(2), 331–338. <https://doi.org/10.1111/j.1467-7687.2009.00888.x>
- Ferguson, H. J., Apperly, I., & Cane, J. E. (2017). Eye tracking reveals the cost of switching between self and other perspectives in a visual perspective-taking task. *Quarterly Journal of Experimental Psychology* (2006), 70(8), 1646–1660. <https://doi.org/10.1080/17470218.2016.1199716>
- Field, A. (2013). *Discovering statistics using IBM SPSS Statistics: And sex and drugs and rock 'n' roll* (4th edition). Sage Publications Ltd.

- Forgács, B., Gervain, J., Parise, E., Csibra, G., Gergely, G., Baross, J., & Király, I. (2020). Electrophysiological investigation of infants' understanding of understanding. *Developmental Cognitive Neuroscience*, 43, 100783. <https://doi.org/10.1016/j.dcn.2020.100783>
- Forgács, B., Parise, E., Csibra, G., Gergely, G., Jacquey, L., & Gervain, J. (2019). Fourteen-month-old infants track the language comprehension of communicative partners. *Developmental Science*, 22(2), e12751. <https://doi.org/10.1111/desc.12751>
- Frith, C. D., & Frith, U. (2008a). Implicit and explicit processes in social cognition. *Neuron*, 60(3), 503–510. <https://doi.org/10.1016/j.neuron.2008.10.032>
- Frith, C. D., & Frith, U. (2008b). Implicit and Explicit Processes in Social Cognition. *Neuron*, 60(3), 503–510. <https://doi.org/10.1016/j.neuron.2008.10.032>
- Frith, C. D., & Frith, U. (2012). Mechanisms of social cognition. *Annual Review of Psychology*, 63, 287–313. <https://doi.org/10.1146/annurev-psych-120710-100449>
- Gallese, V. (2007). Before and below 'theory of mind': Embodied simulation and the neural correlates of social cognition. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 362(1480), 659–669. <https://doi.org/10.1098/rstb.2006.2002>

- Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences*, 2(12), 493–501. [https://doi.org/10.1016/S1364-6613\(98\)01262-5](https://doi.org/10.1016/S1364-6613(98)01262-5)
- Garrod, S., & Pickering, M. J. (2004). Why is conversation so easy? *Trends in Cognitive Sciences*, 8(1), 8–11.
- Gazzola, V., Rizzolatti, G., Wicker, B., & Keysers, C. (2007). The anthropomorphic brain: The mirror neuron system responds to human and robotic actions. *NeuroImage*, 35(4), 1674–1684. <https://doi.org/10.1016/j.neuroimage.2007.02.003>
- Gernsbacher, M. A., & Yergeau, M. (2019). Empirical Failures of the Claim That Autistic People Lack a Theory of Mind. *Archives of Scientific Psychology*, 7(1), 102–118. <https://doi.org/10.1037/arc0000067>
- Goldman, A. I. (2006). *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading*. Oxford University Press. <https://www.oxfordscholarship.com/view/10.1093/0195138929.001.0001/acprof-9780195138924>
- Goodman, Y. M. G. and K. S. (1990). Vygotsky in a Whole Language Perspective. In *Making Sense of Learners Making Sense of Written Language*. Routledge.
- Graham, S. A., Sedivy, J., & Khu, M. (2014). That's not what you said earlier: Preschoolers expect partners to be referentially consistent. *Journal of Child Language*, 41(1), 34–50. <https://doi.org/10.1017/S0305000912000530>

- Green, M. F., Horan, W. P., & Lee, J. (2015). Social cognition in schizophrenia. *Nature Reviews Neuroscience*, 16(10), 620–631. <https://doi.org/10.1038/nrn4005>
- Griffiths, D., & Tipper, S. P. (2009). Priming of reach trajectory when observing actions: Hand-centred effects. *Quarterly Journal of Experimental Psychology*, 62(12), 2450–2470. <https://doi.org/10.1080/17470210903103059>
- Groppe, D. M., Urbach, T. P., & Kutas, M. (2011). Mass univariate analysis of event-related brain potentials/fields I: A critical tutorial review. *Psychophysiology*, 48(12), 1711–1725. <https://doi.org/10.1111/j.1469-8986.2011.01273.x>
- Grzadzinski, R., Huerta, M., & Lord, C. (2013). DSM-5 and autism spectrum disorders (ASDs): An opportunity for identifying ASD subtypes. *Molecular Autism*, 4(1), 12. <https://doi.org/10.1186/2040-2392-4-12>
- Guellai, B., & Streri, A. (2011). Cues for Early Social Skills: Direct Gaze Modulates Newborns' Recognition of Talking Faces. *PLOS ONE*, 6(4), e18610. <https://doi.org/10.1371/journal.pone.0018610>
- Hahne, A., Eckstein, K., & Friederici, A. D. (2004). Brain Signatures of Syntactic and Semantic Processes during Children's Language Development. *Journal of Cognitive Neuroscience*, 16(7), 1302–1318. <https://doi.org/10.1162/0898929041920504>

- Hamilton, A. F. de C., & Lind, F. (2016). Audience effects: What can they tell us about social neuroscience, theory of mind and autism? *Culture and Brain*, 4(2), 159–177. <https://doi.org/10.1007/s40167-016-0044-5>
- Hanna, J. E., Tanenhaus, M. K., & Trueswell, J. C. (2003). The effects of common ground and perspective on domains of referential interpretation. *Journal of Memory and Language*, 49(1), 43–61. [https://doi.org/10.1016/S0749-596X\(03\)00022-6](https://doi.org/10.1016/S0749-596X(03)00022-6)
- Happé, F., & Frith, U. (2014). Annual research review: Towards a developmental neuroscience of atypical social cognition. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 55(6), 553–557.
- Happé, F. G. (1994). An advanced test of theory of mind: 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.
- Hasson, U., & Frith, C. D. (2016). Mirroring and beyond: Coupled dynamics as a generalized framework for modelling social interactions. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 371(1693). <https://doi.org/10.1098/rstb.2015.0366>
- Heyes, C. (2014). Submentalizing: I Am Not Really Reading Your Mind. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, 9(2), 131–143. <https://doi.org/10.1177/1745691613518076>

- Holcomb, P. J., Coffey, S. A., & Neville, H. J. (1992). Visual and auditory sentence processing: A developmental analysis using event-related brain potentials. *Developmental Neuropsychology*, 8(2–3), 203–241. <https://doi.org/10.1080/87565649209540525>
- Holt, R. J., Chura, L. R., Lai, M.-C., Suckling, J., von dem Hagen, E., Calder, A. J., Bullmore, E. T., Baron-Cohen, S., & Spencer, M. D. (2014). 'Reading the Mind in the Eyes': An fMRI study of adolescents with autism and their siblings. *Psychological Medicine*, 44(15), 3215–3227. <https://doi.org/10.1017/S0033291714000233>
- Horton, W. S., & Gerrig, R. J. (2005). The impact of memory demands on audience design during language production. *Cognition*, 96(2), 127–142. <https://doi.org/10.1016/j.cognition.2004.07.001>
- Huber, D. E., & O'Reilly, R. C. (2003). Persistence and accommodation in short-term priming and other perceptual paradigms: Temporal segregation through synaptic depression. *Cognitive Science*, 27(3), 403–430. https://doi.org/10.1207/s15516709cog2703_4
- Jacob, L. P. L., & Huber, D. E. (2020). Neural Habituation Enhances Novelty Detection: An EEG Study of Rapidly Presented Words. *Computational Brain & Behavior*, 3(2), 208–227. <https://doi.org/10.1007/s42113-019-00071-w>
- Jouravlev, O., Schwartz, R., Ayyash, D., Mineroff, Z., Gibson, E., & Fedorenko, E. (2019). Tracking Colisteners' Knowledge States During Language

- Comprehension. *Psychological Science*, 30(1), 3–19.
<https://doi.org/10.1177/0956797618807674>
- Juottonen, K., Revonsuo, A., & Lang, H. (1996). Dissimilar age influences on two ERP waveforms (LPC and N400) reflecting semantic context effect. *Cognitive Brain Research*, 4(2), 99–107. [https://doi.org/10.1016/0926-6410\(96\)00022-5](https://doi.org/10.1016/0926-6410(96)00022-5)
- 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 theory of mind use in adults. *Cognition*, 89(1), 25–41.
- Krauss, R. M., & Glucksberg, S. (1977). Social and nonsocial speech. *Scientific American*, 236(2), 100–105.
<https://doi.org/10.1038/scientificamerican0277-100>
- Kronmüller, E., & Barr, D. J. (2007). Perspective-free pragmatics: Broken precedents and the recovery-from-preemption hypothesis. *Journal of Memory and Language*, 56(3), 436–455.
<https://doi.org/10.1016/j.jml.2006.05.002>
- Kulke, L., von Duhn, B., Schneider, D., & Rakoczy, H. (2018). Is Implicit Theory of Mind a Real and Robust Phenomenon? Results From a Systematic Replication Study. *Psychological Science*, 29(6), 888–900.
<https://doi.org/10.1177/0956797617747090>

- Kuperberg, G. R., Delaney-Busch, N., Fanucci, K., & Blackford, T. (2018). Priming production: Neural evidence for enhanced automatic semantic activity preceding language production in schizophrenia. *NeuroImage: Clinical*, 18, 74–85. <https://doi.org/10.1016/j.nicl.2017.12.026>
- Kurby, C. A., & Zacks, J. M. (2008). Segmentation in the perception and memory of events. *Trends in Cognitive Sciences*, 12(2), 72–79. <https://doi.org/10.1016/j.tics.2007.11.004>
- Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event related brain potential (ERP). *Annual Review of Psychology*, 62, 621–647. <https://doi.org/10.1146/annurev.psych.093008.131123>
- Larson, M. J., & Carbine, K. A. (2017). Sample size calculations in human electrophysiology (EEG and ERP) studies: A systematic review and recommendations for increased rigor. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology*, 111, 33–41. <https://doi.org/10.1016/j.ijpsycho.2016.06.015>
- Leekam, S. R., & Perner, J. (1991). Does the autistic child have a metarepresentational deficit? *Cognition*, 40(3), 203–218. [https://doi.org/10.1016/0010-0277\(91\)90025-Y](https://doi.org/10.1016/0010-0277(91)90025-Y)
- Lemon, J. M., Gargaro, B., Enticott, P. G., & Rinehart, N. J. (2011). Executive functioning in autism spectrum disorders: A gender comparison of

- response inhibition. *Journal of Autism and Developmental Disorders*, 41(3), 352–356. <https://doi.org/10.1007/s10803-010-1039-2>
- Leslie, A. M., & Frith, U. (1988). Autistic children's understanding of seeing, knowing and believing. *British Journal of Developmental Psychology*, 6(4), 315–324. <https://doi.org/10.1111/j.2044-835X.1988.tb01104.x>
- Lewis, P. A., Birch, A., Hall, A., & Dunbar, R. I. M. (2017). Higher order intentionality tasks are cognitively more demanding. *Social Cognitive and Affective Neuroscience*, 12(7), 1063–1071. <https://doi.org/10.1093/scan/nsx034>
- Lidzba, K., Schwilling, E., Grodd, W., Krägeloh-Mann, I., & Wilke, M. (2011). Language comprehension vs. language production: Age effects on fMRI activation. *Brain and Language*, 119(1), 6–15. <https://doi.org/10.1016/j.bandl.2011.02.003>
- Lin, S., Keysar, B., & Epley, N. (2010). Reflexively mindblind: Using theory of mind to interpret behavior requires effortful attention. *Journal of Experimental Social Psychology - J EXP SOC PSYCHOL*, 46, 551–556. <https://doi.org/10.1016/j.jesp.2009.12.019>
- Loucas, T., Charman, T., Pickles, A., Simonoff, E., Chandler, S., Meldrum, D., & Baird, G. (2008). Autistic symptomatology and language ability in autism spectrum disorder and specific language impairment. *Journal of Child Psychology and Psychiatry*, 49(11), 1184–1192. <https://doi.org/10.1111/j.1469-7610.2008.01951.x>

- Loukusa, S., Mäkinen, L., Kuusikko-Gauffin, S., Ebeling, H., & Moilanen, I. (2014). Theory of mind and emotion recognition skills in children with specific language impairment, autism spectrum disorder and typical development: Group differences and connection to knowledge of grammatical morphology, word-finding abilities and verbal working memory. *International Journal of Language & Communication Disorders*, 49(4), 498–507. <https://doi.org/10.1111/1460-6984.12091>
- Lövdén, M., & Johansson, M. (2003). Are covert verbal responses mediating false implicit memory? *Psychonomic Bulletin & Review*, 10(3), 724–729. <https://doi.org/10.3758/BF03196538>
- Luck, S. J. (2014). *An introduction to the Event-Related Potential technique* (2nd ed.). MIT Press.
- M Schaller, U., & Rauh, R. (2017). What Difference Does It Make? Implicit, Explicit and Complex Social Cognition in Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders*, 47(4), 961–979. <https://doi.org/10.1007/s10803-016-3008-x>
- MacWhinney, B. (1992). The Competition Model and Foreign Language Acquisition. In M. L. Swartz & M. Yazdani (Eds.), *Intelligent Tutoring Systems for Foreign Language Learning* (pp. 39–50). Springer. https://doi.org/10.1007/978-3-642-77202-3_4
- Maddox, B. B., Brodtkin, E. S., Calkins, M. E., Shea, K., Mullan, K., Hostager, J., Mandell, D. S., & Miller, J. S. (2017). The Accuracy of the ADOS-2 in Identifying Autism among Adults with Complex Psychiatric

- Conditions. *Journal of Autism and Developmental Disorders*, 47(9), 2703–2709. <https://doi.org/10.1007/s10803-017-3188-z>
- Manly, B. F. J. (1997). *Randomization, Bootstrap, and Monte Carlo Methods in Biology* (2nd ed.). London: Chapman & Hall.
- Mantegna, F., Hintz, F., Ostarek, M., Alday, P. M., & Huettig, F. (2019). Distinguishing integration and prediction accounts of ERP N400 modulations in language processing through experimental design. *Neuropsychologia*, 134, 107199. <https://doi.org/10.1016/j.neuropsychologia.2019.107199>
- Mason, R. A., & Just, M. A. (2009). The Role of the Theory-of-Mind Cortical Network in the Comprehension of Narratives. *Language and Linguistics Compass*, 3(1), 157–174. <https://doi.org/10.1111/j.1749-818X.2008.00122.x>
- Matthews, D., Lieven, E., & Tomasello, M. (2010). What's in a manner of speaking? Children's sensitivity to partner-specific referential precedents. *Developmental Psychology*, 46(4), 749–760. <https://doi.org/10.1037/a0019657>
- McIntyre, N. S., Oswald, T. M., Solari, E. J., Zajic, M. C., Lerro, L. E., Hughes, C., Devine, R. T., & Mundy, P. C. (2018). Social cognition and Reading comprehension in children and adolescents with autism spectrum disorders or typical development. *Research in Autism Spectrum Disorders*, 54, 9–20. <https://doi.org/10.1016/j.rasd.2018.06.004>

- McRae, K., & Matsuki, K. (2013). Constraint-based models of sentence processing. In *Sentence Processing*. Psychology Press.
- Melton, A. W., & Lackum, W. J. V. (1941). Retroactive and proactive inhibition in retention: Evidence for a two-factor theory of retroactive inhibition. *The American Journal of Psychology*, 54, 157–173.
<https://doi.org/10.2307/1416789>
- Milton, D. (2012). On the ontological status of autism: The ‘double empathy problem’. *Disability & Society*, 27(6), 883–887.
- Moreno, E. M., Casado, P., & Martín-Loeches, M. (2016). Tell me sweet little lies: An event-related potentials study on the processing of social lies. *Cognitive, Affective & Behavioral Neuroscience*, 16(4), 616–625.
<https://doi.org/10.3758/s13415-016-0418-3>
- Nadig, A. S., & Sedivy, J. C. (2002). Evidence of Perspective-Taking Constraints in Children’s On-Line Reference Resolution. *Psychological Science*, 13(4), 329–336. <https://doi.org/10.1111/j.0956-7976.2002.00460.x>
- Naughtin, C. K., Horne, K., Schneider, D., Venini, D., York, A., & Dux, P. E. (2017). Do implicit and explicit belief processing share neural substrates? *Human Brain Mapping*, 38(9), 4760–4772.
<https://doi.org/10.1002/hbm.23700>
- Nieuwland, M. S., & Van Berkum, J. J. A. (2006). When Peanuts Fall in Love: N400 Evidence for the Power of Discourse. *Journal of Cognitive*

Neuroscience, 18(7), 1098–1111.

<https://doi.org/10.1162/jocn.2006.18.7.1098>

Paunov, A. M., Blank, I. A., & Fedorenko, E. (2019). Functionally distinct language and Theory of Mind networks are synchronized at rest and during language comprehension. *Journal of Neurophysiology*, 121(4), 1244–1265. <https://doi.org/10.1152/jn.00619.2018>

Perner, J., & Leekam, S. (2008). The Curious Incident of the Photo that was Accused of Being False: Issues of Domain Specificity in Development, Autism, and Brain Imaging. *Quarterly Journal of Experimental Psychology*, 61(1), 76–89. <https://doi.org/10.1080/17470210701508756>

Pfeifer, J. H., & Blakemore, S.-J. (2012). Adolescent social cognitive and affective neuroscience: Past, present, and future. *Social Cognitive and Affective Neuroscience*, 7(1), 1–10. <https://doi.org/10.1093/scan/nsr099>

Radvansky, G. A., Krawietz, S. A., & Tamplin, A. K. (2011). Walking through doorways causes forgetting: Further explorations. *The Quarterly Journal of Experimental Psychology*, 64(8), 1632–1645. <https://doi.org/10.1080/17470218.2011.571267>

Ramenzoni, V. C., Riley, M. A., Shockley, K., & Davis, T. (2008). Carrying the height of the world on your ankles: Encumbering observers reduces estimates of how high an actor can jump. *Quarterly Journal of*

- Experimental Psychology* (2006), 61(10), 1487–1495.
<https://doi.org/10.1080/17470210802100073>
- Rieth, C. A., & Huber, D. E. (2017). Comparing different kinds of words and word-word relations to test an habituation model of priming. *Cognitive Psychology*, 95, 79–104.
<https://doi.org/10.1016/j.cogpsych.2017.04.002>
- Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, 27, 169–192.
<https://doi.org/10.1146/annurev.neuro.27.070203.144230>
- Rueschemeyer, S.-A., Gardner, T., & Stoner, C. (2015). The Social N400 effect: How the presence of other listeners affects language comprehension. *Psychonomic Bulletin & Review*, 22(1), 128–134.
<https://doi.org/10.3758/s13423-014-0654-x>
- 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.
<https://doi.org/10.1037/a0018729>
- Santiesteban, I., Catmur, C., Hopkins, S. C., Bird, G., & Heyes, C. (2014). Avatars and arrows: Implicit mentalizing or domain-general processing? *Journal of Experimental Psychology. Human Perception and Performance*, 40(3), 929–937. <https://doi.org/10.1037/a0035175>

- Santiesteban, I., Shah, P., White, S., Bird, G., & Heyes, C. (2015). Mentalizing or submentalizing in a communication task? Evidence from autism and a camera control. *Psychonomic Bulletin & Review*, 22(3), 844–849. <https://doi.org/10.3758/s13423-014-0716-0>
- Scheeren, A. M., de Rosnay, M., Koot, H. M., & Begeer, S. (2013). Rethinking theory of mind in high-functioning autism spectrum disorder. *Journal of Child Psychology and Psychiatry*, 54(6), 628–635. <https://doi.org/10.1111/jcpp.12007>
- Schneider, D., Bayliss, A. P., Becker, S. I., & Dux, P. E. (2012). Eye movements reveal sustained implicit processing of others' mental states. *Journal of Experimental Psychology. General*, 141(3), 433–438. <https://doi.org/10.1037/a0025458>
- Schneider, D., Lam, R., Bayliss, A. P., & Dux, P. E. (2012). Cognitive load disrupts implicit theory-of-mind processing. *Psychological Science*, 23(8), 842–847. <https://doi.org/10.1177/0956797612439070>
- Schneider, D., Slaughter, V. P., Bayliss, A. P., & Dux, P. E. (2013). A temporally sustained implicit theory of mind deficit in autism spectrum disorders. *Cognition*, 129(2), 410–417. <https://doi.org/10.1016/j.cognition.2013.08.004>
- Schurz, M., Kronbichler, M., Weissengruber, S., Surtees, A., Samson, D., & Perner, J. (2015). Clarifying the role of theory of mind areas during visual perspective taking: Issues of spontaneity and domain-specificity.

NeuroImage, 117, 386–396.

<https://doi.org/10.1016/j.neuroimage.2015.04.031>

- Schuwerk, T., Vuori, M., & Sodian, B. (2015). Implicit and explicit Theory of Mind 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>
- Scott, R. M., & Baillargeon, R. (2017). Early False-Belief Understanding. *Trends in Cognitive Sciences*, 21(4), 237–249. <https://doi.org/10.1016/j.tics.2017.01.012>
- Scott, R. M., & Roby, E. (2015). Processing Demands Impact 3-Year-Olds' Performance in a Spontaneous-Response Task: New Evidence for the Processing-Load Account of Early False-Belief Understanding. *PloS One*, 10(11), e0142405. <https://doi.org/10.1371/journal.pone.0142405>
- Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: Bodies and minds moving together. *Trends in Cognitive Sciences*, 10(2), 70–76. <https://doi.org/10.1016/j.tics.2005.12.009>
- Sebanz, N., Knoblich, G., & Prinz, W. (2003). Representing others' actions: Just like one's own? *Cognition*, 88(3), B11–B21. [https://doi.org/10.1016/S0010-0277\(03\)00043-X](https://doi.org/10.1016/S0010-0277(03)00043-X)
- Sebanz, N., Knoblich, G., & Prinz, W. (2005). How two share a task: Corepresenting stimulus-response mappings. *Journal of Experimental Psychology. Human Perception and Performance*, 31(6), 1234–1246. <https://doi.org/10.1037/0096-1523.31.6.1234>

- Senju, A. (2012). Spontaneous theory of mind and its absence in autism spectrum disorders. *The Neuroscientist: A Review Journal Bringing Neurobiology, Neurology and Psychiatry*, 18(2), 108–113.
<https://doi.org/10.1177/1073858410397208>
- Senju, A. (2013). Atypical development of spontaneous social cognition in autism spectrum disorders. *Brain & Development*, 35(2), 96–101.
<https://doi.org/10.1016/j.braindev.2012.08.002>
- Senju, A., Southgate, V., Miura, Y., Matsui, T., Hasegawa, T., Tojo, Y., Osanai, H., & Csibra, G. (2010). Absence of spontaneous action anticipation by false belief attribution in children with autism spectrum disorder. *Development and Psychopathology*, 22(2), 353–360.
<https://doi.org/10.1017/S0954579410000106>
- Senju, A., Southgate, V., White, S., & Frith, U. (2009). Mindblind Eyes: An Absence of Spontaneous Theory of Mind in Asperger Syndrome. *Science*, 325(5942), 883–885.
<https://doi.org/10.1126/science.1176170>
- Siegal, M., & Varley, R. (2002). Neural systems involved in ‘theory of mind’. *Nature Reviews Neuroscience*, 3(6), 463–471.
<https://doi.org/10.1038/hnrn844>
- Silva, C., Da Fonseca, D., Esteves, F., & Deruelle, C. (2017). Seeing the funny side of things: Humour processing in Autism Spectrum Disorders. *Research in Autism Spectrum Disorders*, 43–44, 8–17.
<https://doi.org/10.1016/j.rasd.2017.09.001>

- Simon, J. R. (1990). The Effects of an Irrelevant Directional CUE on Human Information Processing. In R. W. Proctor & T. G. Reeve (Eds.), *Advances in Psychology* (Vol. 65, pp. 31–86). North-Holland.
[https://doi.org/10.1016/S0166-4115\(08\)61218-2](https://doi.org/10.1016/S0166-4115(08)61218-2)
- Skipper, J. I. (2015). The NOLB model: A model of the natural organization of language and the brain. In R. M. Willems (Ed.), *Cognitive Neuroscience of Natural Language Use* (pp. 101–134). Cambridge University Press.
<https://doi.org/10.1017/CBO9781107323667.006>
- Skipper, J. I., & Lametti, D. R. (2021). Speech Perception under the Tent: A Domain-general Predictive Role for the Cerebellum. *Journal of Cognitive Neuroscience*, 33(8), 1517–1534.
https://doi.org/10.1162/jocn_a_01729
- Šoškić, A., Jovanović, V., Styles, S. J., Kappenman, E. S., & Ković, V. (2021). How to do Better N400 Studies: Reproducibility, Consistency and Adherence to Research Standards in the Existing Literature. *Neuropsychology Review*. <https://doi.org/10.1007/s11065-021-09513-4>
- Southgate, V., Senju, A., & Csibra, G. (2007). Action anticipation through attribution of false belief by 2-year-olds. *Psychological Science*, 18(7), 587–592. <https://doi.org/10.1111/j.1467-9280.2007.01944.x>
- Southgate, V., & Vernetti, A. (2014). Belief-based action prediction in preverbal infants. *Cognition*, 130(1), 1–10.
<https://doi.org/10.1016/j.cognition.2013.08.008>

- Spivey, M. J., & Tanenhaus, M. K. (1998). Syntactic ambiguity resolution in discourse: Modeling the effects of referential context and lexical frequency. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 24(6), 1521–1543. <https://doi.org/10.1037//0278-7393.24.6.1521>
- Steele, V. R., Bernat, E. M., Broek, P. van den, Collins, P. F., Patrick, C. J., & Marsolek, C. J. (2013). Separable Processes Before, During, and After the N400 Elicited by Previously Inferred and New Information: Evidence from Time-Frequency Decompositions. *Brain Research*, 1492C, 92. <https://doi.org/10.1016/j.brainres.2012.11.016>
- Steffensen, S. C., Ohran, A. J., Shipp, D. N., Hales, K., Stobbs, S. H., & Fleming, D. E. (2008). Gender-selective effects of the P300 and N400 components of the visual evoked potential. *Vision Research*, 48(7), 917–925. <https://doi.org/10.1016/j.visres.2008.01.005>
- Surtees, A., Apperly, I., & Samson, D. (2016). I've got your number: Spontaneous perspective-taking in an interactive task. *Cognition*, 150, 43–52. <https://doi.org/10.1016/j.cognition.2016.01.014>
- Swallow, K. M., Zacks, J. M., & Abrams, R. A. (2009). Event Boundaries in Perception Affect Memory Encoding and Updating. *Journal of Experimental Psychology. General*, 138(2), 236. <https://doi.org/10.1037/a0015631>

- Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*, 185(4157), 1124–1131.
<https://doi.org/10.1126/science.185.4157.1124>
- Van Berkum, J. J. A., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating Upcoming Words in Discourse: Evidence From ERPs and Reading Times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(3), 443–467.
<https://doi.org/10.1037/0278-7393.31.3.443>
- Van Berkum, J. J. A., van den Brink, D., Tesink, C. M. J. Y., Kos, M., & Hagoort, P. (2007a). The Neural Integration of Speaker and Message. *Journal of Cognitive Neuroscience*, 20(4), 580–591.
<https://doi.org/10.1162/jocn.2008.20054>
- Van Berkum, J. J. A., van den Brink, D., Tesink, C. M. J. Y., Kos, M., & Hagoort, P. (2007b). The Neural Integration of Speaker and Message. *Journal of Cognitive Neuroscience*, 20(4), 580–591.
<https://doi.org/10.1162/jocn.2008.20054>
- Van Overwalle, F., & Vandekerckhove, M. (2013). Implicit and explicit social mentalizing: Dual processes driven by a shared neural network. *Frontiers in Human Neuroscience*, 7, 560.
<https://doi.org/10.3389/fnhum.2013.00560>
- Vavatzanidis, N. K., Mürbe, D., Friederici, A. D., & Hahne, A. (2018). Establishing a mental lexicon with cochlear implants: An ERP study with

young children. *Scientific Reports*, 8(1), 910.
<https://doi.org/10.1038/s41598-017-18852-3>

- Vesper, C., Abramova, E., Bütepage, J., Ciardo, F., Crossey, B., Effenberg, A., Hristova, D., Karlinsky, A., McEllin, L., Nijssen, S. R. R., Schmitz, L., & Wahn, B. (2017). Joint Action: Mental Representations, Shared Information and General Mechanisms for Coordinating with Others. *Frontiers in Psychology*, 7. <https://doi.org/10.3389/fpsyg.2016.02039>
- Wang, S., Yang, C., Liu, Y., Shao, Z., & Jackson, T. (2017). Early and late stage processing abnormalities in autism spectrum disorders: An ERP study. *PLoS ONE*, 12(5). <https://doi.org/10.1371/journal.pone.0178542>
- Wardlow, L. (2013). Individual Differences in Speaker's Perspective Taking: The Roles of Executive Control and Working Memory. *Psychonomic Bulletin & Review*, 20(4). <https://doi.org/10.3758/s13423-013-0396-1>
- Warnell, K. R., & Redcay, E. (2019). Minimal coherence among varied theory of mind measures in childhood and adulthood. *Cognition*, 191, 103997. <https://doi.org/10.1016/j.cognition.2019.06.009>
- Wechsler, D. (2008). *Wechsler Adult Intelligence Scale—Fourth Edition technical and interpretive manual*. Pearson.
- Westley, A., Kohút, Z., & Rueschemeyer, S.-A. (2017). 'I know something you don't know': Discourse and social context effects on the N400 in adolescents. *Journal of Experimental Child Psychology*, 164, 45–54. <https://doi.org/10.1016/j.jecp.2017.06.016>

- Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, 13(1), 103–128. [https://doi.org/10.1016/0010-0277\(83\)90004-5](https://doi.org/10.1016/0010-0277(83)90004-5)
- Zhang, Q., Lawson, A., Guo, C., & Jiang, Y. (2006). Electrophysiological correlates of visual affective priming. *Brain Research Bulletin*, 71(1–3), 316–323. <https://doi.org/10.1016/j.brainresbull.2006.09.023>