At the Emergence of Theory of Mind: Novel Paradigms for Mental States Language and False Belief Reasoning

Dissertation
zur Erwerb des Doctor of Philosophy (Ph.D) am
Munich Center of the Learning Sciences
der Ludwig-Maximilians-Universität München

vorgelegt von

Stella Serena Grosso

aus

Treviso

2023

At the Emergence of Theory of Mind: Novel Paradigms for Mental States Language and False Belief Reasoning

Stella Serena Grosso

Doctoral Thesis for the Promotion to
Doctor of Philosophy (Ph.D.)
at the Munich Center of the Learning Sciences
of the Ludwig–Maximilians–Universität
München

submitted by
Stella Serena Grosso

Munich, November 11th, 2022

Referent/in: Prof. Dr. Paul Sauseng

Korreferent/in: PD Dr. Tobias Schuwerk

Tag der mündlichen Prüfung: 23.03.2023

Contents

List of Figures **vii**

List of Tables **ix**

Abstract **xi**

1 Main Introduction **1**
- 1.1 Conceptual Foundation . . . 1
  - 1.1.1 Theory of Mind: a Universal Ability . . . 1
  - 1.1.2 The Emergence of Theory of Mind . . . 4
  - 1.1.3 Neural and Cognitive Basis of Theory of Mind . . . 5
  - 1.1.4 Development of Neural and Cognitive Processes of Theory of Mind 7
- 1.2 Problem Statement and Previous Research . . . 9
  - 1.2.1 Explicit False Belief Reasoning . . . 11
  - 1.2.2 Implicit False Belief Reasoning . . . 13
  - 1.2.3 Language Abilities and Theory of Mind . . . 15
- 1.3 Main Objective and Research Questions . . . 18
- 1.4 Dissertation Structure . . . 20

2 Study A: Explicit False Belief Reasoning **23**

3 Study B: Implicit False Belief Reasoning **29**
- 3.1 Abstract . . . 29
- 3.2 Introduction . . . 30
- 3.3 Methods . . . 39
  - 3.3.1 Participants . . . 39
  - 3.3.2 Experimental Paradigm . . . 40
  - 3.3.3 Procedure . . . 42
  - 3.3.4 EEG Recording and ERP Processing . . . 43
  - 3.3.5 Data Analysis . . . 45
- 3.4 Results . . . 46
  - 3.4.1 Behavioural Results . . . 46
  - 3.4.2 ERP Results . . . 47

3.5 Discussion 51

4 Study C: Mental States Language Comprehension 55

4.1 Abstract 55
4.2 Introduction 56
4.3 Methods 60
4.3.1 Participants 60
4.3.2 Experimental Paradigm 61
4.3.3 Procedure 63
4.3.4 Data Pre-processing 66
4.3.5 Data Analysis 68
4.4 Results 70
4.4.1 Behavioural Results 70
4.4.2 Eye-Tracking Results: Test Phase 1 71
4.4.3 Eye-Tracking Results: Test Phase 2 74
4.4.4 Eye-Tracking Results: Test Phase 3 76
4.4.5 Correlations with Mental States Production 78
4.5 Discussion 78

5 Main Discussion 85

5.1 Main Summary 85
5.2 Main Findings 86
5.3 Theoretical and Practical Implications 88
5.3.1 Explicit False Belief Reasoning and Inhibitory Control 88
5.3.2 Processing Demands in Implicit False Belief Tasks 91
5.3.3 The Interplay of Language and Theory of Mind 94
5.3.4 Developmental and Clinical Implications 97
5.4 Limitations and Future Research 100
5.5 Conclusion 105

Bibliography 109

Acknowledgments 129

List of Figures

3.1 Explicit False Belief EEG Tasks 41
3.2 Grand-Average ERP Waveforms 48
3.3 Topographic Heatmap of Scalp Electrical Activity 49

4.1 Mental States Language Comprehension Task 63
4.2 Time course of PTL to Salient Agents in T1 and T2 73
4.3 Time course of PTL to Salient Locations in T3 77

List of Tables

3.1 Summary of Mean Area Amplitudes at 250-350ms 49
3.2 Summary of Mean Area Amplitudes at 650-800ms 50

4.1 Summary of Mean Proportional Looking Times 71
4.2 Reduced General Linear Mixed-Model for Test Phase 1 72
4.3 Full General Linear Mixed-Model for Test Phase 1 72
4.4 Reduced General Linear Mixed-Model for Test Phase 2 75
4.5 Full General Linear Mixed-Model for Test Phase 2 75
4.6 General Linear Mixed-Model for Test Phase 3 77

Abstract

Theory of Mind refers to the cognitive ability of reasoning over others' points of views, goals, desires, and, ultimately, understanding and inferring their mental states.

Recent research in neuropsychology challenged the assumption that Theory of Mind develops only after the fourth year of age, supporting the alternative claim that positive performance in younger population is possible, but it is concealed due to the high cognitive requirements caused by classic paradigms. It is hence necessary to reconsider and revisit traditional Theory of Mind tasks by application of suitable methodologies that can reduce cognitive load.

The primary aim of this project is to confirm the robustness of this claim and to accordingly develop novel paradigms with reduced working memory load and inhibitory control demands, in order to tackle two cognitive mechanisms connected to the development of Theory of Mind: understanding of false belief and comprehension of mental states language.

To achieve this goal, three experimental studies have been implemented here. In the first study, a replication of a behavioural explicit false belief task with a larger sample of 33-month-old children was performed. The replication of this explicit task, characterised by low inhibitory requirements, served as theoretical basis for the development of the following paradigms. We successfully replicated the original findings, showing significant high performance in explicit false belief in 3-year-old children when inhibitory demands are reduced.

Secondly, implicit false belief understanding was investigated in an adult sample with a novel EEG paradigm with reduced processing demands. Neural ERP correlates of implicit false belief discrimination were evoked at distinct latencies. This indicates the novel task effectiveness in detecting false belief reasoning, even when belief-related inferential demands are removed.

Thirdly, an eye-tracking task to explore spontaneous correlates of early comprehension of mental states verbs was developed and validated with a representative sample of 27-month-old children. Analysis of proportional looking time indicated an early sensitivity to epistemic language discrimination already in 2.5-year-old children, with a stunning preference for the novel verb *think* vs. the familiar verb *know*.

Taken all together, this project supports the conceptual continuity hypothesis, which states that false belief discrimination capacities are already present before the fourth year of age, but are obscured by competing executive demands. Furthermore, the development towards a full-fledged Theory of Mind could be facilitated by early mental states comprehension already in the third year of life. This opens the path to future investigations, which can apply these paradigms in longitudinal studies to explore developmental patterns. As a final point, this works confirms the necessity of cross-disciplinary methodologies to be employed synergistically in infant research to investigate effectively the mechanisms behind the emergence of Theory of Mind.

Keywords: Theory of Mind, false belief, ERP, mental states language, eye-tracking, cognitive processing, executive functions, inhibitory control, replication

Chapter 1

Main Introduction

1.1 Conceptual Foundation

1.1.1 Theory of Mind: a Universal Ability

On April 8^{th}, 1300, during his divine journey through the underworld, Dante descends into the fifth circle of hell. There he is moved by the tragic story of Paolo and Francesca, violently killed for their forbidden love:

...And I began: "Thine agonies, Francesca,
Sad and compassionate to weeping make me.
But tell me, at the time of those sweet sighs,
By what and in what manner Love conceded,
That you ***should know*** *your* ***dubious desires****?"*
And she to me: "There is no greater sorrow
Than ***to be mindful*** *of the happy time*
In misery, and that thy Teacher ***knows****.*

Dante, Inferno, V

Great literature speaks to our soul about the most intimate processes of human experience. To understand how Francesca fell into sin, Dante asks her to express when she became aware of her own wishes. In return she states her suffering in recalling past emotions during her eternal punishment. In order to demonstrate the deep compassion that Dante feels for Francesca, he makes us aware of her perspective, emotions, goals, and desires by conveying her mental states. Thinking of another in terms of thoughts, imagining, wanting, and knowing is one of the most fundamental aspects of human life (Wellman, 2014).

As social beings, this mentalising processing can often happen even without verbal communication with fellow humans. Consider the following scenario: you are waiting for the next bus at the stop, together with a stranger. They frantically tap their foot on the ground, a frown on their face, often checking the time on a wristwatch. All of a sudden, they take their phone out of the pocket and - after what seems to be a short but very intense browsing moment - they walk away. What would you do in this situation? Would you just wait for the bus, or would you check the timetable?

It seems clear that the stranger is at the stop to take a bus (as you are), and their behaviour suggests that something is wrong, most probably (we can speculate) that the bus is late. However, this seemingly trivial deduction involves a very non-trivial process: inferring the aims and the beliefs of a complete stranger from their actions.

This cognitive mechanism is the main focus of Theory of Mind. This psychological concept refers to our ability to attribute independent mental states (such as beliefs, desires, goals, wishes, and intentions) to ourselves and others in order to predict and explain their behaviour. Since mental states are not directly observable, the complex system of inferences needed to understand others' behaviours and make predictions about their future actions is referred to as a theory (Premack & Woodruff, 1978).

As illustrated in the two examples above, Theory of Mind allows us to infer information about the world based on others' behaviours and mental states. Theory of Mind is employed in every task involving human interaction, such as cooperative behaviours and social actions, which address directly or indirectly other people (see Azarian, 2010 for a review on social dynamics). It is worth mentioning that Theory of Mind focusses on the mentalising processes of cognitive mental states, relevant when judging, analysing, and inferring others' actions, as well as inferring emotions and emotional experiences of another (so called *cognitive* empathy, see Jolliffe & Farrington, 2004), rather than actually experiencing emotional states of others (so called *emotional* empathy).

We can confidently say that possessing a functional Theory of Mind is crucial for our success in all human social interactions. Many psychological and behavioural disorders have in fact been characterized by concurrent deficits in Theory of Mind, such as autism (Baron-Cohen, 2000 for a review; Frith, 1994), attention deficit hyperactivity disorder (Maoz et al., 2019; Pineda-Alhucema et al., 2018), and schizophrenia (Brüne, 2005; Sprong et al., 2007). More controversial is the impairment of Theory of Mind within the realm of personality disorders, with positive evidence from studies on borderline (Németh et al., 2018) and dark triad personality constructs (Ali & Chamorro-Premuzic, 2010), but negative for clinical psychopathy (Blair et al., 1996; Dolan & Fullam, 2004).

Many factors showed to contribute to the ability of one person to exhibit Theory of Mind, among which especially age, cognitive delays, and language acquisition are of particular interest for neuropsychological research. In fact, Theory of Mind develops in childhood in parallel with empathy, linguistic skills, executive functions, and other higher order cognitive abilities (Carlson et al., 2002; Carlson et al., 2004; Hughes & Ensor, 2007; Levine, 2009).

It is not surprising then that our current understanding of the discovery and development of Theory of Mind comes mainly from developmental psychology research on infants. How Theory of Mind develops is not yet fully understood, but studying the emergence of Theory of Mind at its early stages is an exciting challenge, with high potential to provide us with the necessary evidence to one day unravel the mystery behind an essential social ability.

1.1.2 The Emergence of Theory of Mind

Until recently, the fourth year of age was considered a milestone in the infant to preschooler development of a full-fledged Theory of Mind (Gómez, 2009; Wimmer & Perner, 1983). At this stage, children are consistently successful in solving classic paradigms aiming to tackle the understanding of others' false beliefs, such as the change of content and the change of location task, where they show competence in predicting the behaviour of others based on their belief (see Wellman et al., 2001, for a meta-analysis). This finding has been attributed to a conceptual change in the way children process and represent mental states and are able to disregard the true (or their own) state of reality in favour of others' possibly flawed representations (see Sabbagh et al., 2010; Wellman, 2011; Wellman & Cross, 2001).

Despite this, in the last fifteen years growing evidence has been brought in favour of Theory of Mind competences developing earlier on in life, when less traditional paradigms are used to test children appreciation of mental states.

One of the first pivotal studies in this regard comes from Clements & Perner (1994), which developed an anticipatory looking paradigm for false belief understanding. In this study, the authors showed how children that are 2 years and 11 months old were consistently able to look in anticipation at agents aiming to the location where they thought their desired object would be versus the place where the object really was. With the same paradigm, Southgate et al. (2007) showed correct anticipation of actions of an agent holding a false belief already in 25-month-old infants recorded with an eye-tracker.

In addition, indication of predictive actions based on an adult's previous beliefs was found in as early as 18-month-old infants with an anticipatory intervening paradigm (Knudsen & Liszkowski, 2012) and with active helping behavioural tasks (Buttelmann et al., 2009).

Lastly, a much discussed outcome was brought to the public by Onishi & Baillargeon (2005), who developed a non-verbal violation of expectation task, where the agents looked for an object in a location they would not have known it would be versus a scene where the agents looked in a location they have seen the object previously being hidden into. In their task, which made use of eye-tracking technology, looking time showed evidence of predictive behaviour towards the location the agents' false belief pointed to already in 15-month-old infants.

Overall, these controversial findings at very early age sparked considerable discussion in the scientific community on the timeline of development and emergence of a reduced versus full-fledged Theory of Mind mechanism in the infant's mind.

1.1.3 Neural and Cognitive Basis of Theory of Mind

It follows that an enormous research interest has spurred in the last decades for investigating the social and biological processing mechanisms underlying the development of this essential ability during childhood (Harris, 2006; Mahy et al., 2014; Wellman, 2002). Nevertheless, while the fourth year of age has been considered the turning point in Theory of Mind maturation, little is known about the neural mechanisms supporting this critical step and how to detangle its developing stages from concurrent cognitive abilities, which are also affected by the development of affected neural regions.

Several cognitive processes are necessary for Theory of Mind reasoning, from building a representation of others' mental states to decoupling between mental states and reality, and inhibiting one's own knowledge to successfully solve the task from the perspective of another agent (Flavell et al., 1990). Neuroimaging studies in adults suggest that the Theory of Mind network involves a number of bilateral regions, including amygdala, anterior cingulate cortex and anterior insula, inferior frontal gyrus, intraparietal and posterior superior temporal sulci (Frith & Frith, 2007; Mar, 2011; van Overwalle, 2009).

An overarching role in reasoning about others' minds has been specifically attributed to the medial prefrontal cortex (mPFC), representing socially relevant information about another person, such as thoughts and beliefs, but also subjective feelings (Saxe & Powell, 2006) and mental adjectives (Mitchell et al., 2006). It is interesting to highlight that variations of Theory of Mind paradigms led to activation of differential networks (Schurz et al., 2014, for a meta-analysis) and, in particular, of the temporoparietal junction (TPJ), a subregion located at the supramarginal gyrus, consistently recruited in false belief understanding tasks. Because of its involvement in the updating of mental states (Aichhorn et al., 2006; Saxe & Kanwisher, 2003; Vogeley et al., 2001), the right TPJ has been considered a candidate cortical area involved in integrating sensory events of contextual relevance for reorienting our representation of the world, creating perspective differences necessary for the decoupling mechanism between mentality and reality at the basis of Theory of Mind processing (Sommer et al., 2007).

Paralleling neuroimaging findings, Electro-Encephalography (EEG) studies in adults supported a possible role of the right TPJ during reasoning about others' mental states, reflected in a right posterior activation during false belief tasks (Geangu et al.,2012; Liu et al., 2009a; Meinhardt et al., 2011). In particular, the last decade saw a rise for studies demonstrating that distinct Event-Related Potentials (ERP) correlates reflect false belief reasoning in our brain. Judgments about another person perspective have been repeatedly connected with the emergence of a late slow wave (LSW) at anterior regions in tasks comparing belief-based and non-mental written representations

(Sabbagh & Taylor, 2000), verbally narrated stories about reasoning about beliefs versus reality (Liu et al., 2004), written characters' attribution stories comparing desires versus beliefs (Liu et al., 2009a), and non-verbal false versus true belief photographic stories (Geangu et al., 2012).

Further studies suggested that it is possible to discriminate an earlier component represented by the late positive complex (LPC), a waveform with similar characteristics of the attention-related P3a component (Polich, 2007). The LPC has been associated with the initial state of identification of another's perspective and mental states categorization (Jiang et al., 2016), and it could be the neural marker of bottom-up cognitive processes antecedent to mental states reasoning, such as reorientation from external to internal mental representations, e.g. when attention from an agent's behaviour shifts to one's own internal knowledge about the agent's beliefs (Meinhardt et al., 2011). Other interpretations support a role of LPC as a correlate of the mental states processing effort required in tasks with high cognitive demands for the integration of information for reasoning (Cao et al., 2012), as well as for inhibitory control of reality or one's own perspective in order to infer another's beliefs (Zhang et al., 2009).

From these sparse and not always consistent findings, it follows that we are still lacking a clear understanding of what is the spatial and temporal neural dimensions of Theory of Mind processing, and particularly of false belief reasoning.

1.1.4 Development of Neural and Cognitive Processes of Theory of Mind

It is therefore little wonder that the literature on the neurophysiological markers of false belief understanding is still scarce and little is known about the development of the neural processes during the emergence of Theory of Mind in early childhood. A pivotal false belief study using Track-Density Imaging methodology recently showed that only after sufficient white matter structure maturation of temporo-parietal and medial frontal areas, such as the connection between TPJ and anterior areas via the arcuate fasciculus, children show a mature explicit representation of mental states resembling the one of the adult population (Grosse Wiesmann et al., 2017). Modulation of TPJ activity was also seen in neuroimaging studies where 8- to 12-year-old children solved verbal and nonverbal second-order false belief tasks (Kobayashi et al., 2007), while 10- to 12-year-old children did not selectively recruited the right TPJ during an unexpected transfer false belief paradigm, in contrast to adults (Sommer et al., 2010).

This age-related maturation regards core areas of the adult Theory of Mind network, such as TPJ and mPFC, which show communalities with recent results coming from developmental EEG studies on false belief reasoning. Resting alpha activity localized in the right TPJ was positively associated with 4-year-olds' performance in representational Theory of Mind tasks (Sabbagh et al., 2009). In ERP research, a diffuse frontal negative LSW was found in 4- to 6-year-old children who were competent in a think judgment task when contrasting false and true belief reasoning with reality judgements, similar to adults (Liu et al., 2009b). Using an EEG-suitable version of the traditional unexpected transfer task for testing 6- to 8-year-old children, Meinhardt et al. (2011) found that false belief reasoning elicited both an early LPC component, interpreted as neural marker of the cognitive processes required for updating the mental model of the environment, and a positive LSW, associated with the conceptual load due to decoupling between false belief and reality. A fronto-central distributed LSW specific to false belief understanding was also found in a more recent study on pretense versus false belief reasoning in 6- to 8-year-old children (Kühn-Popp et al., 2013).

All these studies considered, there are still many questions and little answers regarding what are the early neural correlates evoked by Theory of Mind scenarios, what they truly represent in terms of cognitive processes, and which neurobiological changes are necessary for the development of adult-like competences in false belief tasks. In this regard, one of the main goal of this dissertation is to provide further insights on the neural correlates of implicit false belief in adults.

1.2 Problem Statement and Previous Research

It is not surprising that behavioural and neurophysiological early age positive findings about false belief reasoning sparked heated debate (see Scott & Baillargeon, 2014, for an overview). Ultimately, evidence of false belief understanding before the age of four gave voice to supporters of an early-developing Theory of Mind hypothesis, which suggests the presence of a specialized neurocognitive mechanism already mature in the second year of life (Leslie, 2005), in open contrast to the developmental Theory of Mind change view (Ruffman & Perner, 2005; Saxe et al., 2004).

In order to explain the positive results obtained before four years of age, traditional supporters of the developmental Theory of Mind change view suggested that these tasks are flawed in their main intent, as children do not necessarily need to attribute belief to others and their task performance could be attributed to the surprising presence of unfamiliar actions (Perner & Ruffman, 2005).

A more optimistic, comprehensive hypothesis instead supports a conceptual continuity view (Baillargeon et al., 2010) between early and late findings, though affirming that young children already possess a concept of belief, which guides their expectations of others' actions. In this view, it is primarily due to excessive language requirements, computational demands, working memory constraints, and lack of inhibitory control that children competences are masked in traditional false belief tasks, where open response is required (Rubio-Fernández & Geurts, 2013). In particular, traditional false belief tasks like the Sally-Anne Scenario (see Chapter 3 for a detailed explanation) require attentional (it is after all an intricate narrative involving two agents) and linguistic resources (such as appreciation of subtle semantic differences, i.e. *Where will Sally look* versus *Where should she look*) which are too hard to process for infants (Bloom & German, 2000).

In support of this hypothesis, there is evidence that 3-year-olds are able to successfully pass simpler variants of the classic false belief task. Infants showed high performance in paradigms where a memory aid for the false belief content was given (German & Leslie, 2000), where the saliency of the change of location event was lowered (Wellman & Bartsch, 1988), or where children were asked to actively choose (instead of predict) the correct location based on the agents' belief, and the desired object was removed from the scene, hence eliminating its saliency (Carpenter et al., 2002).

Taken all together, it seems reasonable to assume that children's performance in false belief tasks is sensitive to what seems minor changes in extraneous tasks demands due to inefficient processing capacities (German & Leslie, 2000). Therefore, classic methodologies to detect and quantitatively estimate early Theory of Mind abilities show limitations whose true explanatory impact we started to appreciate only recently. New paradigms should take advantage of the combined expertise from different disciplines to disentangle the tight relation between higher order cognitive development, Theory of Mind emergence, and language acquisition. These combined multidisciplinary approaches can be valid in tackling Theory of Mind questions in early childhood without the technical pitfalls of traditional paradigms.

In this dissertation, we aimed to validate current low-demand approaches and foster new approaches, which are suitable to test children younger than three years of age, merging methods from developmental psychology with expertise coming from neuroscience and linguistics research. In particular, we decided to focus on two of the most important cognitive resources involved in the development of a full-fledged Theory of Mind: explicit and implicit false belief reasoning and language comprehension of mental states.

1.2.1 Explicit False Belief Reasoning

In the previous section, we mentioned how false belief understanding is traditionally considered the litmus paper of Theory of Mind abilities. Explicit false belief refers to the ability of children to give overt, elicited response about their understanding of others' false beliefs during a task involving their active participation (Grosse Wiesmann et al., 2017).

After all, the inability for children younger than four years of age to pass explicit false belief task could be due to their inadequacy in representational resources to pass this task. As previously proposed, an alternative explanation is that methodological problems are behind the capacity of younger children to show their true understanding of false belief.

Firstly, classic explicit task have intrinsic limitations due to cognitive demands and memory requirements (Carpenter et al., 2002). Although behavioural false belief studies are best developed, results have been often confounded by the presence of paradigms features that can be considered disadvantages (e.g. the necessity to elicit verbal response) and do not clearly allow detangling the domain-specificity of the cognitive effect under examination.

A second methodological problem regarding classic false belief studies concerns the replicability of previous findings. These failed replications have been explained in terms of procedural differences between original and replication studies, as well as by attention and motivation differences (Baillargeon et al., 2018), which led to massive exclusion of subjects following original study criteria (Schuwerk et al., 2018).

We are well aware that these fragile experimental studies are at the basis of all theoretical conceptualization of false belief reasoning in childhood. In the spirit of the contemporary replication efforts in psychology (Simmons et al., 2016), we decided to perform a replication attempt focussing on the striking difference in performance of children younger than four years of age between traditional and simpler, low-demands false belief tasks. Following the main drive to investigate the applicability and robustness of explicit false belief tasks with different cognitive demands, we developed a behavioural replication of an explicit false belief task with low inhibitory demands (Setoh et al., 2016) within a large scale sample of 33-month-old children. Furthermore, we aimed to compare children behavioural performance in this paradigm with a second yet traditional representational change false belief task, which has previously showed competence in children younger than 3 years of age (Hughes & Ensor, 2007).

We selected this recent study by Setoh et al. (2016) because of its peculiarity and yet simplicity in addressing the issue of high inhibitory control demands of the traditional false belief task. In this paradigm, two main variations from the traditional false belief task are implemented: firstly, the desired object was removed from the scene. The rationale behind this choice is that inappropriate response due to the current state of knowledge of the child should be in such way suppressed. Secondly, to further reduce cognitive demands, the task included two practice questions that helped the child to get familiar with the goal of the game and know what to expect next. We targeted our replication effort on a sample of 33-month-old children, in order to investigate whether we can find evidence of false belief understanding even at such an early age, when concurrent inhibitory and response generation demands are addressed to reduce cognitive processing requirements.

1.2.2 Implicit False Belief Reasoning

Behavioural developmental research found multiple evidence of early sensitivity for mental states related to Theory of Mind in children before four years of age. Already in their second year of life, infants can track agent's perspective and encode agent-event relations relevant to false belief reasoning (Rakoczy, 2012; Sodian, 2011), as well as show behaviours that require attribution of mental sates to others, such as imitate intended actions (Carpenter et al., 1998) or adapting their behaviour based on the knowledge of the state of the world of their parents (O'Neill, 1996).

Hence, we have evidence of infants primordial skills in mentalising others' actions, a view in agreement with the emergence of early implicit false belief reasoning: children spontaneous understanding of others' (false) mental states (Baillargeon et al., 2010), elicited by implicit performance markers that can be quantitatively measured, such as anticipatory and preferential looking times or neurophysiological activation patterns. As presented in the previous sections, spontaneous response tasks showed positive results earlier than the four years of age threshold, with an exception for verbal tasks, which often require significant linguistic skills, too demanding for toddlers and comparable to the ones of explicit false belief tasks. Recently though, evidence has been brought that even 2.5-year-olds can pass verbal tasks when not asked to elicit a direct response, thus supporting the continuity view of early false belief understanding (Scott et al., 2012).

At this point, addressing the elephant in the room is essential: the debate around the presence of an early implicit false belief reasoning mechanism has been possible thanks to the employment of techniques that do not require elicited answer from the participant. Powerful tools for investigating the implicit understanding of Theory of Mind at its emergence have been Eye-Tracking (see studies on the above-mentioned claim that false belief reasoning can be elicited in preverbal infants if cognitive demands are reduced, e.g. Baillargeon et al., 2010; Southgate et al., 2007) as well as

Electro-Encephalography (EEG), functional Near-Infrared Spectroscopy (fNIRS), and functional Magnetic Resonance Imaging (fMRI).

In the last two decades, there has been a considerable rise in studies demonstrating that distinct Event-Related Potentials (ERPs) can be identified as neural marker of Theory of Mind abilities. ERP studies in the Theory of Mind domain focus on temporal aspects of the decoding, processing, and integration of mental states such as desires, emotions, and especially beliefs. In particular, specific ERPs have been correlated to false belief reasoning, often via the employment of EEG-adapted versions of traditional behavioural false belief tasks. As presented in the previous section on the neural and cognitive basis of Theory of Mind, these studies have brought new heat to the early sensitivity debate, offering meaningful insights into the early cognitive processing of mental states.

To date, we have no knowledge of ERP studies tackling the cognitive development of false belief understanding below the age of four, at the emergence of Theory of Mind ability. Although the potential of these methods is clear, this could be due to the several challenges rising from employing neurophysiological techniques on young children. On one side, even though EEG is a completely non-invasive and not-interfering method for gathering neurophysiological information, the relative discomfort of the cap and cables needed for recording and the high number of trial repetitions is a challenge for the motivation and attention span of infants, ultimately leading to low signal-to-noise ratio due to high movement artefacts. On the other side, these studies often suffer from similar drawbacks of the traditional explicit false belief tasks, such as requiring high memory workload due to long story formats. It follows that a difficulty in studying the neurophysiological basis of Theory of Mind in childhood is to disentangle genuine neurocognitive components of Theory of Mind reasoning from other rather general neurocognitive processes, which are also required for solving the task.

Among these processes a special role is played by executive functions, i.e. the complex set of mechanisms which allows to focus on information and shift attention effectively (e.g. working memory and selective attention), inhibit impulsive responses (i.e. inhibitory control), generate strategic decision-making, flexible problem solving, and goal-direct behaviour (Diamond, 2013). A conclusion derived from spontaneous response tasks is that children younger than four years of age already have an understanding of false beliefs, but not yet sufficiently developed inhibitory control, which hinders correct overt responses (Scott & Baillargeon, 2009; for an alternative explanation see Perner & Ruffman, 2005). Further support to this hypothesis comes from recent behavioural research, showing that when reducing executive functions load, children as young as 33 months are able to pass a traditional Theory of Mind task (Setoh et al., 2016, supported by Study A in this dissertation).

Therefore, our goal as researchers should be to consider prior findings from behavioural, facilitated false belief studies focussing on lower cognitive requirements and combine these with previous research on the neural correlates of false belief. Hence, here we aim to develop a new false belief task with reduced processing demands, where the spontaneous response, as measured by the ERP event, is time-locked to the agent eliciting its own mental states in agreement (true) or disagreement with reality (false belief). The study includes the developing of a full new paradigm based on a neurophysiologically compatible, digital version of a behavioural task by Wellman & Bartsch (1988): to test its validity in eliciting neural correlates, a sample of thirty neurotypical adults has been included in this pivotal phase of the paradigm.

1.2.3 Language Abilities and Theory of Mind

The emergence of a full-fledged Theory of Mind seems to be the supported by several developmental factors, such as the evolution of social competences, the refinement of executive functioning skills, and especially the acquisition of language abilities. Correlations between proficiency in language skills and Theory of Mind has been found in 3- to 6-year-old children (Harris et al., 2005), to the point of suggesting that language could be a necessary precursor to successful development of a Theory of Mind (Astington & Jenkins, 1999; Ruffman et al., 2002). In particular, it has been argued that earlier general language abilities drives later explicit false belief performance, based on a longitudinal meta-analysis (Milligan et al., 2007). More controversially, false belief understanding has been also proposed to be dependent on the mastery of syntactic complement structures (De Villiers & De Villiers, 2000), with evidence of predictive directionality between complement usage and false belief performance already in 3- to 4-year-old children (De Villiers & Pyers, 2002).

These results are by some means in agreement with the hypothesis that the third year of life plays an important role for the transition from implicit to explicit understanding of verbal Theory of Mind (Diessel & Tomasello, 2001). Despite this, little evidence has been found to correlate implicit false belief understanding with language (Low, 2010; Grosse Wiesmann et al., 2017). One of the reason for this could be the challenge to test language comprehension concurrently with implicit false belief reasoning before four years of age.

Mental States Comprehension

Of particular interest for us is the less contentious hypothesis that it is the acquisition of a verbal system of mental states representation (e.g. epistemic states, desires, and perceptions) to be essential for fostering children's conceptual understanding of false belief (Harris et al., 2005). In principle, it is known that at 2- to 3-year-old children produce their first mental verbs and verbs of communication (Diessel & Tomasello, 2001). This is in line with the results from a longitudinal study from Brook &

Meltzoff (2015), where they found evidence of the bridging role of language and non-social attention in the development of infants' social cognition first and Theory of Mind later. In particular, children with higher gaze-following behaviour at 10.5 months showed higher performance in mental states words production at 2.5 years of age, and the same held true for mental states words production at 2.5 years of age and higher success in standard explicit Theory of Mind tasks at 4.5 years of age. This predictive relationship was specific for mental states representations, in agreement with a new study showing that children's mental states language are concurrently related with Theory of Mind independently of general language abilities, and only training with mental states language promotes acquisition of understanding of knowledge and ignorance already at 33 months of age (Kaltefleiter et al., 2022). Nevertheless, although initial production of mental states starts early in infancy, full mastery of the semantics of mental states terms only occurs around 4 to 5 years of age (Moore et al., 1989), leading to theorize that the acquisition of mental states language could be the true precursor to a full-fledged Theory of Mind (Kristen-Antonow et al., 2019).

However, verbal production does not necessarily mean that implicit understanding has to develop simultaneously. A first discrimination in implicit mental states language before the age of four would be meaningful, in light of recent studies showing that 2-year-olds already show epistemic state production in spontaneous speech, although hardly talking about third parties (Harris et al., 2017).

To date, early comprehension of mental states language in children below three years of age has not been systematically investigated. One of the main challenges is yet a methodological one, since the sparse evidence brought from behavioural paradigms for testing mental states language comprehension show poor performance in children below four years of age (Moore et al., 1989). So far, mental states language competence has been mainly assessed by parent report questionnaires and analysis of spontaneous speech samples (such as from the CHILDES database, see MacWhinney, 2000). As a consequence, this is a limiting factor for investigating how early mental states skills link to implicit false belief understanding.

To ultimately reach this goal, it is necessary first to develop a more sensitive task to tackle mental states comprehension early on, without relying on elicited production from children. For this purpose, we developed an eye-tracking study in close collaboration with Prof. Nivedita Mani (Department of Psychology of Language, Georg-August-Universität Göttingen) to test spontaneous responses, as elicited by a proportional looking paradigm, in discriminating between the semantics of the mental states verbs *know* and *think* in 27-month-old children. If successful, such task has the potential to be employed as part of a language battery to investigate the relation between implicit false belief understanding and early access to mental states semantics before the fourth year of age.

1.3 Main Objective and Research Questions

Leitmotif of this work is the necessity for methodological improvements of traditional tasks, which are at the basis of theoretical conceptualization of, and multiple debates on, Theory of Mind mechanisms in early childhood. Treasuring the experience acquired from traditional false belief tasks, we built the foundations for a theory that sees the fourth year of age as a milestone for the development of Theory of Mind. Yet, recent experimental research continuously brings us surprising insights on the child's mind before this magical year has been reached. Here, we want to combine teachings from the limitations of classic tasks and the subsequent criticism that led to innovative and yet debated paradigms, to remedy flaws and improve the robustness of our current understanding.

Our goal is to test and develop novel study paradigms combining traditional expertise from developmental psychology with spontaneous-response approaches used in neurophysiological studies and linguistic research, to ultimately shed further light onto two of the most important cognitive resources necessary for the development of Theory of Mind in infancy: the understanding of false belief and the comprehension of mental states language.

The following research questions have been tackled in this work, divided in three experimental studies:

A) Can we replicate with a larger sample the positive findings from an explicit false belief behavioural task in children below the age of four, when cognitive demands are reduced?

To answer this question, in Study A we replicated with a statistically sounded sample size a behavioural explicit false belief study that uses a paradigm with reduced information processing demands (Setoh et al., 2016). This study hypotheses are that 1. Children as early as 33-month-of-age have an appreciation for false belief understanding, 2. Inhibitory control demands play an impactful role in performance obtained in explicit tasks, and 3. Positive findings from behavioural studies on early false belief can be replicated with a larger sample size.

B) Can we develop a child-friendly EEG task with reduced processing demands that tackles implicit false belief understanding at the neural level?

In study B we developed a novel ERP study suitable for young children to investigate the neural correlates of implicit false belief understanding. In order to do so, we implemented the learnings from Study A) on the necessity to reduce extraneous cognitive load and adapted the behavioural paradigm from Wellman & Bartsch (1988), where extraneous cognitive requirements were reduced by removing inferential reasoning in favour of elicited statements on the agent's belief. This study hypotheses are that 1. A reduced processing demands ERP task elicits consistent markers of false belief understanding in a pilot sample of neurotypical adults participants, and 2. Such markers should be compatible with previous neurophysiological literature of false belief comprehension (i.e. LPC- and LSW-like waveforms).

C) Can we develop a child-friendly eye-tracking task that measure implicit sensitivity to semantic concepts at an early age? Do children as early as 27 months of age already show initial discrimination for mental states language?

Finally, in study C we developed a new paradigm taking advantage of proportional looking times to investigate comprehension of mental states language in young children, adapting the classic behavioural paradigm developed by Moore et al. (1989). This study hypotheses are that 1. A spontaneous-response looking time paradigm can be used to elicit markers of sensitivity to semantic discriminations, 2. Such sensitivity is already visible in children below 3 years of age, and 3. Children sensitivity should show a looking preference towards the agent holding the highest degree of knowledge.

1.4 Dissertation Structure

In this dissertation, a chapter is dedicated to each of the following three experimental studies after the Main Introduction (Chapter 1).

Study A (Chapter 2 in this manuscript) appeared as Article published in Infant Behavior and Development, Volume 54, Stella S. Grosso, Tobias Schuwerk, Larissa J. Kaltefleiter, Beate Sodian, *33-month-old children succeed in a false belief task with reduced processing demands: A replication of Setoh et al. (2016)*, Pages 151-155, Copyright Elsevier (2019). ISSN 0163-6383, https://doi.org/10.1016/j.infbeh.2018.09.012.

Study B (Chapter 3: Implicit False Belief Reasoning with Reduced Processing Demands: Evidence from Neural Correlates) and Study C (Chapter 4: 27-month-olds'

Early Sensitivity to Mental States Language Comprehension: An Eye-tracking Study) follow a journal article structure with abstract, introduction, methods, results, and discussion sections.

The dissertation closes with a Main Discussion (Chapter 5), Bibliography, and Acknowledgements.

These studies have been funded by the German Research Foundation (DFG grant SO 213/33-1) as part of the Research Group ”Crossing the borders: The interplay of language, cognition, and the brain in early human development (FOR 2253). Project Part 8: The role of language in early Theory of Mind development.”

Declarations of Interest: None.

Chapter 2

Study A: Explicit False Belief Reasoning

33-month-old children succeed in a false belief task with reduced processing demands: A replication of Setoh et al. (2016)

Infant Behavior and Development 54 (2019) 151–155

Contents lists available at ScienceDirect

Infant Behavior and Development

journal homepage: www.elsevier.com/locate/inbede

33-month-old children succeed in a false belief task with reduced processing demands: A replication of Setoh et al. (2016)

Stella S. Grosso*, Tobias Schuwerk, Larissa J. Kaltefleiter, Beate Sodian

Department of Psychology, Ludwig Maximilian University, Munich, Germany

ARTICLE INFO

Keywords:
Theory of mind
False belief understanding
Explicit task
Replication
Low-inhibition
Representational change

ABSTRACT

A recent low-inhibition false belief task showed a high success rate with 33-month-old children when response-generation demands were reduced [Setoh, Scott, & Baillargeon (2016). *Proceedings of the National Academy of Sciences*, *113*(47), 13360–13365]. We found correct responding in 74% of N = 58 33-month-old children, replicating the original findings. Within the same sample, we compared this performance with performance in a concurrent measure of false belief understanding which has previously produced competence in children below the age of 3 years [Hughes & Ensor (2007). *Developmental Psychology*, *43*(6), 1447–1459]. Contrasting sharply with findings from the low-inhibition false belief task, we found partial competence in 15%, and full competence in only 5% of the same sample. These results show that the paradigm by Setoh and colleagues generates reliable findings in a different lab and a different language. We discuss this pattern of results in relation to theoretical considerations of early false belief understanding.

1. Introduction

Theory of Mind describes our ability to impute mental states, such as desires or beliefs, to agents to explain and predict their behavior. Becoming able to correctly explain and predict behavior based on false beliefs, i.e. beliefs that do not correspond to the state of reality, is considered a milestone in Theory of Mind development. In traditional false belief tasks that ask for a verbal prediction of a false belief-based action, children show this ability at around four years of age (Wimmer & Perner, 1983).

Recently, Setoh, Scott, and Baillargeon, (2016) published first evidence that toddlers as young as 2.5 years are able to verbally predict a false belief-based action in a modification of a traditional false belief task. This finding of an early false belief competence is in line with evidence from spontaneous response tasks that draw on gaze patterns or neural responses to indicate false belief processing even in infancy (Southgate, Senju, & Csibra, 2007; Onishi & Baillargeon, 2005; Southgate & Vernetti, 2014; see Sodian, 2016, for an overview). Yet, because Setoh et al. show this competence in verbal responses, which leave less interpretational ambiguity as compared to the more indirect measures in spontaneous response tasks, their study is of particular importance.

Setoh et al. (2016) presented children with a story-format task that introduced an agent who placed an object in location A, which was subsequently moved to location B, unbeknown to the agent. Over 70% of the 30- to 33-month old children correctly indicated the location where the agent was going to search for the object, based on her false belief. This finding contrasts sharply with the findings of hundreds of studies (for a meta-analysis, see Wellman, Cross, & Watson, 2001) showing that children below the age of 4 years indicate the wrong location (the location where the object really is) in such tasks.

Based on an expanded processing account of children's performance on false belief tasks, Setoh et al. (2016) argued that young

* Corresponding author.
E-mail address: stella.grosso@psy.lmu.de (S.S. Grosso).

https://doi.org/10.1016/j.infbeh.2018.09.012
Received 29 April 2018; Received in revised form 2 August 2018; Accepted 28 September 2018
Available online 26 October 2018
0163-6383/

children may fail traditional false belief tasks because of overwhelming concurrent processing demands in the task. During a traditional false belief task, the child is not only required to represent the false belief of the agent, but also to suppress the prepotent response to answer where the object actually is and to generate the correct response. Limited cognitive resources to process the latter two were suggested as possible causes preventing successful task performance and thereby masking false belief understanding. To lower the processing demands, Setoh et al. designed a traditional false-belief task in which inhibitory control demands are reduced by removing the critical object from the scene. In addition, they promoted response generation via the insertion of two practice questions. In these practice questions, the children were asked – just as in the false belief test question – about an object's location. The two practice trials supposedly helped the children in three ways, namely making them acquainted with "where" questions, allowing them to choose between two pictures and select one of them, and giving them clues about when a question is going to be asked.

In one experiment, over 70% of both 30- and 33-month-old children succeeded in this task. In a second experiment, where practice trials differed from the test trial in one aspect, i.e. involving a pointing response only to one instead of two pictures, only 33-month-olds succeeded. Further experiments showed that one practice question instead of two was not enough to lower the response generation demands, and that just lowering the response generation demands was not successful when the inhibition requirements of the task were high. These findings support the authors' conclusion that toddlers possess a conceptual understanding of false belief that is continuous with belief understanding in infancy (Scott & Baillargeon, 2017), but that they are unable to express their understanding in standard tasks due to overwhelming processing demands. It has been argued, however, that in Setoh et al.'s task the practice trials, training children to respond to "where" questions, may have prompted them to point to the last location where the object was, thus leading to success on the task without attributing a false belief to the agent (Rubio-Fernández, Jara-Ettinger, & Gibson, 2017; but see Scott, Setoh, & Baillargeon, 2017, for a reply).

The purpose of the present study was to replicate the findings by Setoh et al. (2016) in a larger sample (cf., Button et al., 2013). Testing the replicability of this modification of a traditional false belief task that relies on verbal responses is even more important in the light of the recently fast-growing body of failed replication attempts of spontaneous response tasks, testing false belief understanding in infancy (e.g., Kulke, Reiß, Krist, & Rakoczy, 2018; Schuwerk, Priewasser, Sodian, & Perner, 2018).

In order to test for the presumed difference in processing demands between Setoh et al.'s low-inhibition false belief task and a traditional task, we included a second false belief task in a within participant design. A representational change false belief task by Hughes and Ensor (2007) has shown success in about a third of 2-year-old children, and thus seems to be sensitive to a nascent false belief understanding in toddlers. In an engaging picture-book format, the child is presented with a display that appears to show an eye of an animal which, on turning the page, is revealed to really be a spot on a snake's skin. Children are asked for both their own prior false belief and the false belief of another person. They are also required to answer two reality control questions. If the Setoh et al. task is a valid measure of false belief understanding, we would expect performance to be correlated with the representational change false belief task despite a sizeable difference in difficulty.

2. Methods

Following best practice recommendations, we report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study (Simmons, Nelson, & Simonsohn, 2012). The ethics committee of the Department of Psychology and Education of Ludwig-Maximilians-University (Munich, Germany) approved this study. All procedures performed in the reported experiments were in accordance with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

2.1. Participants

A total of 62 children took part in the experiment. Informed consent was obtained by one of the parents before starting the experimental session. We applied the same exclusion criteria as described in the original studies, leading to slightly varying sample sizes for each task. In this replication attempt of the low-inhibition false belief task by Setoh et al. (2016), the final sample included 58 children (35 females) with a mean age of 33 months and 7 days (*SD* = 7.6 days, age range = 32 months and 15 days – 34 months). Another four children were tested but not included in the final sample because of failing to point (n = 2), or responding ambiguously in the test trial (the verbal response and pointing gesture did not match, n = 2). In our replication attempt of the representational change false belief task by Hughes and Ensor (2007), 48 children (27 females) with a mean age of 33 months and 6 days (*SD* = 7.9 days, age range = 32 months and 15 days – 34 months) were included. Fourteen children were a priori excluded from analyses due to failing to respond to any of the questions (n = 6), experimental errors (n = 5), not wanting to do the task (n = 1), or receiving help from caretakers (n = 2). Due to failing in replying during either control or test questions or both, sample size further varied between analyses.

The present study was part of a longitudinal study. Therefore, sample sizes were a priori determined based on the design of this larger study. Simonsohn (2015) recently suggested that the sample size of a replication study should be 2.5 times larger than the sample size of the original study in order to protect original findings from underpowered replication attempts. Based on this criterion, our sample of n = 58 for the replication of Setoh et al. (2016) was sufficiently large (sixteen 33-month-olds in the original study). However, our sample (n = 42) for the replication of Hughes and Ensor (2007) is lower than in the original study (n = 122).

Both tasks were part of a test battery of an ongoing longitudinal study. Due to the complexity of the longitudinal design, we refrained from introducing an additional factor by counterbalancing the order of presentation of the two current false belief tasks. They were administered in one test session with a fixed order, intermixed with further tasks not relevant for this study. The children completed the representational change false belief task first. Because (1) materials of the two false belief tasks substantially differed, (2) they were separated by other unrelated tasks, and (3) they were of different types (representational change vs. change of location), we argue that potential carry-over effects are unlikely.

2.2. Low-inhibition false belief task by Setoh et al. (2016)

2.2.1. Material

For the low-inhibition false belief task adapted from Setoh et al. (2016), we created a picture book with nine pages composed of clear plastic sheet protectors (32 cm × 56 cm) holding white paper backgrounds to which 11 picture photos were attached (20 cm × 25 cm). A black solid paperboard stand (50 cm tall × 56 cm wide × 20 cm deep) kept the pages in place via four binder rings mounted on the upper edge. The stand allowed the screen to be positioned at a 70° angle via a black ribbon connecting both the frontal and back frames. All photos were centered at the bottom of the page, with double photos placed 4.5 cm apart one from the other. The pictures created for this task adaption can be found here: < https://osf.io/aejz6/?view_only=e688b18b4f9a40b58b5f7671022423c8 > .

2.2.2. Coding

The child and experimenter sat next to each other and the stand was placed in the middle of the table in front of the child. An overhead camera captured this setting in top view. A second camera, positioned behind the child, captured the experimenter, the book and the child´s responses. A second experimenter coded the child's behavior during the test session. All responses were independently coded from video recordings by other two raters, who reached very high inter-rater reliability for all questions (Cohen's *kappa* = 1). For each practice and test trial we coded the pointing and/or vocal reply of the child and eventual specific response behavior (e.g., inconsistent verbal/pointing responses, i.e. pointing at one location but mentioning the other). Two children required a second prompt during the practice trials.

2.2.3. Procedure

In total, six story events, two practice trials, and one test trial were presented to the child. After flipping each page to reveal the respective picture(s), the experimenter recited the accompanying line of that event. For this replication, the original phrases were translated into German. The script and stimuli material can be found at the link given before. In the first two story events, the protagonist, Lily, found an apple in a bucket covered with a towel. In the first practice trial, a picture of the apple and a picture of a banana were shown and the child was asked "Where is Lily´s apple?". Then, in the third and fourth story event, Lily moved the apple in a basket covered with a plate and went outside to play with a ball. In the second practice trial, the experimenter presented a picture of a rattle and a picture of the ball and asked "Where is Lily´s ball?". The story continued with the arrival of Lily´s brother Peter, who took away the apple from the scene. In the last story event, Lily came back and looked for her apple. In the test trial, the experimenter revealed a picture of the basket and a picture of the bucket and asked "Where will Lily look for her apple?". A pointing gesture or verbal referral to the container where Lily falsely believed that the apple was located was coded as correct response. Cases in which the child needed prompts (e.g. "Can you show me where Lily's apple is?" or "Show me where the apple is!"), were documented, as in the original study design. Differing from the procedure of the original study, we did not counterbalance the positions of the pictures showing the containers in the test trial (i.e., vary between participants whether the belief-congruent container was left/right) and kept only the one of the two possible picture combinations in the test trial in which container location matched the previous pictures. Our rationale here was again to keep the number of additional factors as low as possible in this longitudinal design. Counterbalancing this and other factors within the administered tasks across participants and measurement points would have exceedingly inflated the overall design of our study.

2.3. Representational change false belief task by Hughes and Ensor (2007)

2.3.1. Material and procedure

In the representational change false belief task adapted from Hughes and Ensor (2007), the child and the experimenter sat together on the same side of the table. The experimenter presented a picture book containing a deceptive element (Moerbeek, 1994) while narrating its story. Each page of the book had a hole revealing a small part of the next page. The part that was visible through this hole looked like an eye. Only on the last page it became clear that what appeared to be an eye was actually a spot on the skin of a snake. After finishing the book, the experimenter returned to the second last page and asked the child the test question "What did you think it was before we turned the page, an eye or a spot?". Subsequently, a reality control question was asked ("What is it really, an eye or a spot?"). After that, a puppet figure called Jonas was introduced to the child. The experimenter explained that Jonas had never seen the content of the book before. Holding the puppet in front of the book, a test question was used to assess understanding of other's belief ("What will Jonas think if we show him this page, is this an eye or a spot?"). The child was then asked a second reality control question as in the first scenario.

2.3.2. Coding

For both test questions, responses were coded as "correct" when the child replied with the word "eye" (one's own and other's previous belief), while for both control questions, responses were coded as "correct" when the word "spot" was mentioned (one's own and other's current belief). A second rater coded the responses offline from video-recordings, reaching a high inter-rater reliability

(Cohen's *kappa* = .96). Following the original coding criteria by Hughes and Ensor (2007), performance was coded as successful only if they answered both the control and the test question correctly. This resulted in a sum score for successfully passing both one's own and the other's belief questions ranging from 0 to 2.

3. Results

In the test trial of the low-inhibition false belief task, 43 out of 58 (74,1%) children chose the container in which Lily falsely believed the apple was located ($p = < .001$, cumulative binomial probability, chance level of 0.5). This replicated the original result reported by Setoh et al. (2016), 33-month-olds: 12 out of 16 correct responses, $p = .038$).

In the representational change false belief task, 2 children out of 42 (4.8%) were able to pass both the one's own belief and the other's belief test questions when applying the criteria by Hughes and Ensor (2007). Analyzing the one's own belief and the other's belief trials separately, we found that 7 out of 46 (15.2%) children passed the own belief question and 4 out of 43 (9.3%) passed the other belief question following the original inclusion criterion of passing also the control question (differences in sample size are due to missing data when combining self-belief and other- belief scores together).

A Wilcoxon Signed-ranks test revealed no significant difference between self- (Mdn = 0) and other- (Mdn = 0) belief scores, suggesting that explicating one's own previous belief is not easier at this age than considering another person's belief, $Z = 1.13$, $p = .257$. The overall performance of our sample in both own and other belief questions ($M = 0.26$, $SD = 0.54$, N = 42) turned out to be considerably lower compared to the findings by Hughes and Ensor (2007) for slightly younger children ($M = 0.65$, $SD = 0.77$, N = 122; Hughes & Ensor, 2007).

Bivariate non-parametric correlation analyses between the low-inhibition false belief task performance and the representational change false belief task scores with original criteria (own belief, n = 44, other belief, n = 41, overall score n = 40) did not yield any significant result (Spearman's rho ranging from - .05 to - .06, *p*s ranging from .696 to .777).

4. Discussion

In this study, we successfully replicated the finding by Setoh et al. (2016) that 33-month-olds succeed in a false belief task with reduced processing demands. In our sample, 74% correctly responded that the agent will look for the object where she falsely believes it would be located. In contrast, we did not replicate Hughes and Ensor (2007) finding in a representational change false belief task, but obtained floor effects. Although we expected performance to be significantly lower than in the task by Setoh et al., we expected about one third of the 33-month-olds to pass the representational change false belief task by Hughes and Ensor (2007) based on these authors' findings. While we do not have an explanation for the discrepancy between our findings and Hughes and Ensor's, it should be noted that other studies of representational change false belief tasks (Gopnik & Astington, 1988) found similar levels of difficulty for this task format as of standard location-change paradigms.

No significant correlation between performance in the two tasks was found. The unexpected poor performance in the representational change false belief task and the consequently reduced variance of performance in this task could be accountable for this result. Further, our samples (varying between 40 and 44 participants for the respective measure) might have been two small to detect any true effects. Yet, if there would have been a strong relationship between the two tasks, which could be expected given that both tasks presumably measure the same underlying phenomenon, our sample would have been large enough to detect this effect. In sum, due to the reduced variance and the relatively small sample size, the results of these correlational analyses must be interpreted with caution.

An isolated evaluation of our successful Setoh et al. (2016) replication supports the authors' interpretation that reducing inhibitory-control demands improves performance in a false belief task in children below the age of 4 and is consistent with the expanded processing account that young children's false belief understanding is masked by not yet sufficiently developed inhibitory control and cognitive resources for response generation. The children's poor performance in the representational change false belief task by Hughes and Ensor (2007) is also in line with this interpretation. This task requires a considerable amount of inhibitory control to suppress one's current state belief in favor of one's and another person's prior false belief, as well as to suppress the response to the previous test question when responding immediately afterwards to the control question.

While the present findings indicate that the paradigm by Setoh et al. (2016) generates reliable findings in different labs and different languages, the present study did not contribute to assessing the validity of the task as a measure of false belief understanding. This was not possible due to the unexpectedly poor performance of 33-month-olds in the representational change false belief paradigm by Hughes and Ensor (2007). Note, however, that the present pattern of findings rules out the interpretation that the high performance obtained by Setoh et al. (2016) could have been due to unusual characteristics of their sample. Further research is needed to determine whether there is a unique source of shared variance between children's responses to the task by Setoh et al. and to traditional false belief tasks. Predictive relations in a longitudinal study would be one promising source of evidence for evaluating the proposal by Setoh et al. (2016) that there is conceptual continuity in false belief understanding from toddlerhood to the age of 5 years.

Declarations of interest

None.

Acknowledgement

This study was funded by the German Research Foundation (DFG grant SO 213/33-1).

References

Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., et al. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience, 14*(5), 365–376.

Gopnik, A., & Astington, J. W. (1988). Children's understanding of representational change and its relation to the understanding of false belief and the appearance-reality distinction. *Child development,* 26–37.

Hughes, C., & Ensor, R. (2007). Executive function and theory of mind: Predictive relations from ages 2 to 4. *Developmental Psychology, 43*(6), 1447–1459.

Kulke, L., Reiß, M., Krist, H., & Rakoczy, H. (2018). How robust are anticipatory looking measures of theory of mind? Replication attempts across the life span. *Cognitive Development, 46*, 97–111.

Moerbeek, K. (1994). *Can't sleep.* London, UK: Western Publishing Company.

Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science, 308*(5719), 255–258.

Rubio-Fernández, P., Jara-Ettinger, J., & Gibson, E. (2017). Can processing demands explain toddlers' performance in false-belief tasks? *Proceedings of the National Academy of Sciences, 114*(19), E3750.

Schuwerk, T., Priewasser, B., Sodian, B., & Perner, J. (2018). The robustness and generalizability of findings on spontaneous false belief sensitivity: A replication attempt. *Royal Society Open Science, 5*(5), 172273.

Scott, R. M., & Baillargeon, R. (2017). Early false-belief understanding. *Trends in Cognitive Sciences, 21*(4), 237–249.

Scott, R. M., Setoh, P., & Baillargeon, R. (2017). Reply to rubio-Fernández et al.: Different traditional false-belief tasks impose different processing demands for toddlers. *Proceedings of the National Academy of Sciences of the United States of America, 114*(19), E3751–E3752.

Setoh, P., Scott, R. M., & Baillargeon, R. (2016). Two-and-a-half-year-olds succeed at a traditional false-belief task with reduced processing demands. *Proceedings of the National Academy of Sciences, 113*(47), 13360–13365.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2012). A 21 word solution. *Dialogue: The Official Newsletter of the Society for Personality and Social Psychology, 26*, 4–7.

Simonsohn, U. (2015). Small telescopes: Detectability and the evaluation of replication results. *Psychological Science, 26*(5), 559–569.

Sodian, B. (2016). Is false belief understanding continuous from infancy to preschool age? In D. Barner, & A. S. Baron (Eds.). *Core knowledge and conceptual change* (pp. 301–321). Oxford: Oxford University Press.

Southgate, V., & Vernetti, A. (2014). Belief-based action prediction in preverbal infants. *Cognition, 130*(1), 1–10.

Southgate, V., Senju, A., & Csibra, G. (2007). Action anticipation through attribution of false belief by 2-year-olds. *Psychological Science, 18*(7), 587–592.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development, 72*(3), 655–684.

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.

Chapter 3

Study B: Implicit False Belief Reasoning with Reduced Processing Demands: Evidence from Neural Correlates

3.1 Abstract

A full-fledged Theory of Mind is considered accessible only around four years of age, when children successfully pass explicit false belief tasks. In the last decades, ERP research investigated the neural correlates of false belief understanding in a variety of EEG-adapted classic paradigms in the child and adult populations. So far, no ERP study focused on tackling implicit false belief understanding before the fourth year of age, at the onset of development of Theory of Mind abilities.

One of the reasons for children to fail false belief tasks earlier on has been attributed to high tasks requirements, such as cognitively demanding inferences necessary to solve traditional false belief tasks. Here we developed a novel ERP paradigm with reduced processing demands, to shed further light onto the neural correlates of implicit false belief reasoning. Due to its simple short-story format and lack of false belief inferential demands, such task would be suitable to assess ERP indicators of false belief reasoning in early years of life, at the time of emergence of false belief understanding.

In the present study, thirty adults were presented with movies in which an agent's false or true belief about an object location was explicitly stated and thus did not have to be inferred. At the neurophysiological level, a positive component for false belief emerged in the left hemisphere at 250-400 ms, resembling the LPC previously found in literature. A later divergence, comparable to the LSW, emerged in frontal regions and with reverse polarity at central-parietal sites between 650-800 ms. These characteristic waves have often been associated to sequential temporal stages of the cognitive processing of Theory of Mind.

These findings support the validity of this new paradigm with reduced processing demands to elicit specific markers associated with belief discrimination and are consistent with previous outcomes on the neural correlates of false belief reasoning. Lastly, these results support the task suitability to investigate implicit false belief in infancy.

Keywords: Theory of Mind, implicit false belief understanding, ERP, reduced processing, LPC, LSW

3.2 Introduction

As part of a social community, processing perspectives and opinions different from our own and interpreting actions of others based on their beliefs, goals, and desires is an important aspect of our life. All these are elements of what constitutes Theory of Mind, the social-cognitive skill necessary to impute states of mind to others and distinguish them from our own. As experience shows us, these individual mental states do not always have a correspondence with the actual state of reality, as ignorance and errors of judgment can lead us to inaccurate or erroneous points of view about the world. False belief reasoning describes our ability to attribute such misconceptions to others and ourselves, in order to explain and predict human behaviour during social interactions.

Belief reasoning has been long considered fundamental for the development of a full-fledged Theory of Mind in children (Wellman et al., 2001; Wimmer & Perner, 1983), yet, its underpinning neural mechanisms are not fully understood. Event-Related Potential (ERP) research evoked by belief reasoning found evidence of early and late neural correlates, interpreted as markers of different cognitive processing stages involved in the unfolding of this complex ability. Two main ERP components seem to be recurrently involved in Theory of Mind processing: a broader, mental states related late positive complex (LPC) and a late slow wave (LSW) specific to belief reasoning. At present, there is mixed evidence in the ERP literature in support of the cognitive interpretation of the LSW and LPC, especially due to the difficulty in detangling their either task- or Theory of Mind-specific nature.

In particular, neurophysiological studies implemented a few similar but not overlapping paradigms, focussing on contrasting the concept of belief with other mental states (e.g. desire and pretense reasoning) as well as with non-mental states (e.g. visual representations), and ultimately specifically comparing true versus false belief (Jiang et al., 2016, for an overview). As ERP correlates are intrinsically connected to the paradigm and presentation format used in the experiment (Handy, 2005; Luck, 2014), there is no agreement yet on an unambiguous interpretation of the cognitive processes associated with the same component, when this is evoked in studies that employ different paradigms.

Notably, mental states decoding studies used a variety of tasks, which brought evidence in support of the involvement of the LPC. Sabbagh et al. (2004) found a posterior P300 – 500 component, evoked only at the onset of pictures presented to participants who were asked to judge whether the depicted eyes displayed the same state expressed by the mental states word presented just before. This LPC-like wave was interpreted as expression of the difficulty of the mental states decoding task at hand also in another study, where it was evoked only at the onset of cartoonish pictures when Theory of Mind attribution was necessary for comprehension and thus it was connected to understanding of mental states (Wang et al., 2010). These studies conceived the presence of an LPC as the neural marker of the complexity of the cognitive processes required to separate mental states reasoning from the true state of affair.

In a more recent study, Cao et al. (2012) instead tackled the subtle difference between simple mental states decoding (judging emotional expression of agents on photographs) and reasoning (predicting which object the agent on the photograph will choose). In this case, the authors interpreted a bilateral LPC, which was evoked only at the onset of presentation of target photographs in the reasoning condition, as marker of the greater mental resources necessary to integrate contextual information in mental states reasoning.

All considered, LPC shows some evidence to generically represent the processing demands involved in mentally decoding and representing someone else's mental state, while holding in mind our own knowledge of the true state of the world.

In addition to the LPC, studies that focused specifically on contrasting mental states found evidence of a second, long-lasting LSW component. When participants were asked to compare written characters' contrasting statements about desires, beliefs, and physical locations of objects, Liu et al. (2009a) showed that only explicit judgments based on belief statements evoked a right-posterior LSW, time-locked to the presentation of the object. With the same paradigm, Bowman et al. (2012) showed a right posterior divergence at 600-800ms for belief reasoning compared to desire reasoning also in 7- to

8-year-old children that correctly solved belief reasoning trials in an adult-like manner. Because of its specificity, the right posterior LSW component was interpreted as evidence of a second-stage of belief processing, in agreement with neuroimaging studies that indicate the right temporo-parietal junction as critical region for updating of mental states in the Theory of Mind network (Aichhorn et al., 2006; Kobayashi et al., 2007; Saxe & Kanwisher, 2003; Vogeley et al., 2001).

In addition, the LSW has not only been considered a neural correlate of belief processing, but also more specifically a marker of the mental operations involved in false belief reasoning: the main evidence for this hypothesis comes from a pivotal ERP study by Sabbagh and Taylor (2000). In this study, the authors presented participants with narratives, where agents hold false beliefs of objects location versus non-mental representations, and asked them to elaborate where the agent would imagine the object to be. ERPs were time-locked to the test question and only the false belief reasoning condition evoked a left frontal LSW. In partial agreement with the above-mentioned LSW studies, the authors interpreted this component as reflection of the cognitive processes required to integrate inconsistent mental representations of others (i.e. false belief about the true location of an object) in the current context (i.e. the request to point to the false location).

Taken together, all the mentioned studies show one common feature: evoking Theory of Mind representation in a controlled setting is problematic, as it requires the establishment of a situational context and a suitable narrative to naturally convey the agents' knowledge of the world. It follows that considerable part of the ERP research on false belief focused on implementing an EEG compatible version of a classic, well-established false belief paradigm from behavioural psychology, the unexpected transfer task (Baron-Cohen et al., 1985). In this paradigm, often referred to as the Sally-Anne scenario, Sally hides an object in a location and leaves the room; while absent, Anne moves the object to a new location. At this point, the participant is asked where Sally will look for the object. Success in solving the Sally-Anne scenario, pointing to the original location of the object instead of the new one, is considered a critical test for the emergence of a full-fledged Theory of Mind (Baron-Cohen et al., 1985).

This paradigm has the advantage to compare reality ("Where is the object really?") versus false belief judgments ("Where does he/she think the object is?") in the same scene. In 2004, Liu et al. animated cartoonish stimuli to present their participants with an ERP version of the unexpected transfer task, time-locked to the final questions and evoking a left frontal LSW peak, more negative for belief judgments. This component was interpreted as neural evidence of the decoupling mechanism distinguishing our own mental representation from the true state of affair.

This led to a series of follow-up studies applying variations of the task: in Zhang et al. (2009), participants were presented with a variant of the unexpected transfer task involving two conditions (standard, Sally put the object in another container; and adapted, Sally puts the object away). In contrast with Liu et al. (2004) findings, this paradigm found evidence of a LPC at the onset of the final question for the standard transfer task, but no evidence of the LSW. The unique presence of an LPC was interpreted as indication that inhibition of one's own perspective (the object's actual location, only available in the standard condition), comes temporally earlier than mental operations of false belief conception (associated with a LSW).

We can speculate that a possible explanation for the absence of the LSW in the Zhang et al. (2009) study could lay in its specific task requirements. In a subsequent unexpected transfer study that evoked a frontal LSW, both in adults and 4- to 6-year-old children that correctly passed the task, Liu et al. (2009b) suggested the LSW to be connected to the length of time necessary for holding actively information in verbal working memory and performing conceptual operations to solve mentalising problems.

Despite this interpretation, an anterior LSW was evoked also in a non-verbal version of the unexpected transfer task, when comparing true and false belief in cartoonish stories, both in adults and with longer latencies in 6- to 8-year-old children (Meinhardt et al., 2011). Thus, more than being specifically related to verbal working memory, LSW could be connected to the conceptual load required by mental states attribution in decoupling situations between individual mental states and reality.

In the same study, Meinhardt et al. (2011) found also a parietal LPC, which was interpreted as evidence of the process of shifting our attention from the context where the Theory of Mind situation occurs to our internal representation of the world (a so-called meta-representation, see Perner, 1991). The earlier latency of the LPC compared to the LSW component supports the hypothesis that the emergence of a meta-representation in situation requiring Theory of Mind must temporally precedes mental states decoupling.

A second traditional false belief task that brought support to the decoupling role of the LSW is the unexpected transfer paradigm by Perner et al. (1994). In an ERP adapted version of this task, an anterior LSW was specifically associated to the onset of false belief reasoning in a verbally narrated cartoon sequence involving a reality, a pretense play, and a false belief condition (Meinhardt et al., 2012). With the same paradigm, Kühn-Popp et al. (2013) elicited a bilateral posterior LSW also in 6- to 8-year-old children that correctly solved the task.

All things considered, a shared feature of the false belief paradigms reported here is that participants were explicitly instructed to judge and perform a Theory of Mind task to which the ERP event was time-locked (e.g. “Where will Sally look for the object?”). A different approach comes from Geangu et al. (2012), who showed that explicit instructions are not necessary to elicit false belief components, by using an ERP-adapted version of the belief discrimination task by Onishi and Baillargeon (2005). In this non-verbal paradigm, participants passively viewed photographic sequence of an agent reaching for an object because of her true belief (the object was in the scene, but hidden behind an occluder) or false belief (the object was not in the scene anymore, as it left the scene while hidden behind the occluder). ERPs were time-locked to the final moment in which the agent reaches for the object behind the occluder. Surprisingly, a parietal LPC was found for the true belief condition, while a central and parietal LSW were specifically evoked by false belief. The presence of both components suggests that monitoring others' beliefs can be elicited also in the absence of explicit instructions.

A study by Zhang et al. (2013) confirmed such conclusion comparing true and false belief to a false sign condition, where an arrow points to a location where the agent is but, once the agent moves to a second location, the arrow keeps pointing to the original one. In such situation, the participant was asked to identify where the agent was based on the sign. To no surprise, a frontal, central and parietal LSW was specifically evoked in the classic false belief reasoning condition, compared to true belief and to the false sign condition.

These studies suggest that the same ERPs can be evoked by different paradigms, such as active behavioural response versus passing view, despite very different task requirements. It follows that more investigation is needed in testing generalizability of ERP findings across paradigms. If different studies tackling belief reasoning are indeed evoking similar neural correlates irrespectively of task-specific requirements, we can claim that findings from ERP research are truly markers of temporally specific cognitive mechanisms of Theory of Mind.

In the present study, we aimed to develop an ERP version of a reduced processing demand false belief task, which has been overlooked until now in the investigation of the neural underpinnings of false belief understanding. We are confident that this task can help to shed further light onto which components are associated with false belief reasoning, when cognitive load is lowered and thus less confounding between competing processing activities ensues. In particular, this paradigm avoids a number of difficulties associated with traditional false belief tasks that can potentially confound the interpretation of neural correlates: inferential reasoning (which can be hindered by inhibitory mechanisms), multi-sequence narratives and change of location (both aspects that load working memory).

Firstly, in our task participants were explicitly told about the agent's false belief. Because of having to infer the content of an agent's belief from situational conditions, traditional false belief tasks pose high cognitive demands on inference processes. It is still unclear whether the neural markers accompanying these processes are primarily due to inferential demands (the domain-general process of deducing another agent's opinion) or to false belief representation per se (the domain-specific processing of holding a faulty state of knowledge against reality). In the reduced processing demands task (see Wellman et al., 2001), participants are openly shown or told about the state of reality, e.g., "the key is in the green box" as well as about the agent's belief, e.g., "Paul believes the key is... in the red box" (false belief) versus "...in the green box" (true belief). In order to solve this task, participants only need to compare the information about the agent's belief with the state of reality and notice the discrepancy in the false belief scenario, without having to infer the agent's false belief from situational conditions, such as the agent's access to information at a given point in time. One might expect such a task to be trivially easy, since the content of the belief can be read off the verbally given information. However, behavioural findings from research on 3- to 5-year-old children suggest that this task is equally difficult to traditional false belief tasks where participants have to infer the protagonist's belief from story information (Wellman & Bartsch, 1988) and performance is consistent with standard false belief tasks.

Secondly, we favoured a limited number of photographic scenarios, timely narrated by brief voiceovers, with the aim of reducing extraneous cognitive load (Arguel & Jamet,

2009) and fostering integration of visual and verbal information in working memory (Mayer & Moreno, 1998; Schüler et al., 2011).

Thirdly, we adapted the classic unexpected transfer task, removing the change of location event that could further weight on working memory and inhibitory processing. In a traditional task, the participant needs to hold in memory throughout the story the original location of the object for correctly predicting the agent's behaviour, while maintaining active their own mental representation of reality. In Liu et al. (2009b) a left frontal LWS was interpreted in relation to the length of time working memory is activated, suggesting belief reasoning partly reflects conceptual operations in verbal working memory recruited to solve mentalising problems. Thus, reducing information to be hold in working memory can help to clarify the true nature of this correlate.

In order to achieve these goals, we presented participants with short movies where an agent held either a false or a true belief about an object location, previously introduced to the participant only. A voiceover describing the epistemic state of the agent acted as salient event for eliciting false belief reasoning. ERPs were time-locked to this onset event, and each movie ended with a predictive question about the future action of the agent.

With the employment of this facilitated task, we aim to shed further light onto the neural markers of false belief processing in adults and their generalizability across paradigms. Rather than reflecting generic cognitive processes required for solving the task, we expect that correlates evoked in our task will help to extend findings on the specific neurocognitive processes associated with false belief reasoning.

In particular, if the mental representation necessary for false belief is not dependent on the way false belief is elicited (via inferential reasoning versus explicitly provided), we expect our task to elicit neural components distinguishing between mental processing steps of false versus true belief similar to what has been reported for other

Theory of Mind paradigms in the ERP literature. An anterior LSW should be elicited by false belief reasoning in adults who correctly performed the task, as possible index of decoupling processes (Geangu et al., 2012; Liu et al., 2004, 2009a; Meinhardt et al., 2011; Sabbagh & Taylor, 2000).

Furthermore, if the LPC reflects updating of one's own mental representations, we should find evidence of this neural marker as previously found at temporo-parietal regions (Meinhardt et al., 2011; Zhang et al., 2009). The presence of these components in a facilitated task would remove interpretational confounds previously attributed to overlapping cognitive mechanisms (e.g. working memory load and inhibitory control) and corroborate evidence that ERP results on false belief understanding are not as task-specific as previously thought.

3.3 Methods

3.3.1 Participants

The final sample consisted of 30 adults (16 females and 14 males), with a mean age of 24.12 years (SD = 5.5) and a range between 18 and 40 years of age. Twenty additional participants were tested but not included in the final sample, because of revealing to be bilinguals (which could have affected interpretation of the task) after the experiment (n = 2), misunderstanding of the task instructions based on final written debriefing (n = 4), not providing at least 25 artefact free trials per condition due to blinks, body movements, and tiredness (n = 13) or at least 9 valid responses to control questions per condition (n = 1). All participants were right-handed, native German speakers with normal or corrected-to-normal vision, and had no previous history of neurological or psychological conditions.

Further exclusion criteria were being under psychotropic medications, presence of dermatitis/skin infections or excessive alcohol consumption in the 24 hours previous to the test session. Participants gave written informed consent at the beginning of the study and received a monetary compensation at the end of the experiment.

The ethics committee of the Faculty of Psychology and Educational Sciences of Ludwig-Maximilians-University (Munich, Germany) approved this study. All procedures performed in the reported experiments were in accordance with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

3.3.2 Experimental Paradigm

Each trial consisted of an 18-second-long movie showing a sequence of static frames (see Figure 3.1). At the beginning of each trial, two containers (a bucket and a box) accompanied by a catchy sound were presented to the participant. The two containers remained on the screen over the whole trial in fixed locations. Then, one of four possible toys (a car, a fish, a cup, or a ball) appeared for 2 seconds inside a container. All toys were counterbalanced for appearance location across trials. A narrator voiceover presented the context to the participant (e.g., “Look! This is Leo’s fish. The fish is in the bucket.”).

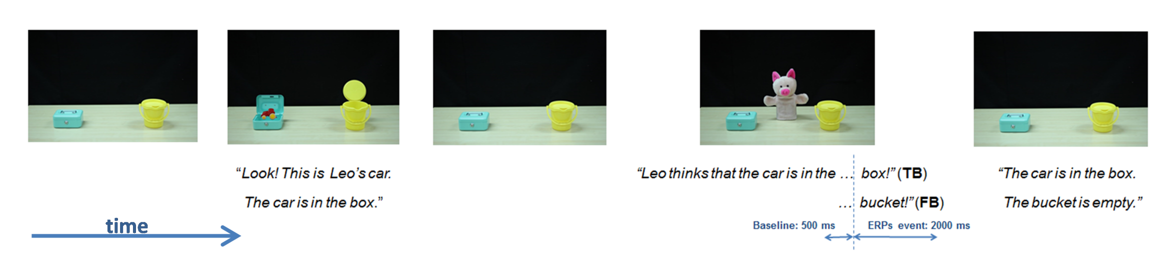
All voiceovers were recorded in German and normalized in voice peak frequency. After disappearance of the object, the agent (Leo) appeared on screen and his belief regarding the object location was verbally stated by the narrator (e.g., “Leo thinks that the fish is in the… box!”). The agent either held a true or a false belief about the location of the object.

ERPs were time-locked at the onset of the object location (the words “bucket” or “box”) following the statement of the agent’s belief. After the agent disappeared, we introduced two filler sentences to avoid biasing the final question with the container location mentioned just before. The first filler was a repetition of the real object’s location (e.g., “The fish is in the bucket.”). The second filler clarified that the other location was empty (e.g., “The box is empty.”). Both fillers were counterbalanced for order across trials. Each trial ended with a test question (e.g., “Where will Leo look for the fish?”) to keep the participant engaged in actively performing the task. The final scene remained on screen until the response was registered by the experimenter. Two versions of control trials where the agent’s belief was substituted with a preference statement for one of the two containers (e.g. “Leo likes… the box!”), counterbalanced for mentioned location, were inserted every 5 test trials.

In the first version, the preference of the agent was in agreement with the current position of the toy (matching condition); in the second version, the agent had an opposite preference to the current toy’s position (mismatching condition). The final control question “Where will Leo go now?” was used to investigate whether participants would choose the preferred location independently of the position of the toy at the behavioural level.

Figure 3.1

Explicit False Belief EEG Task



Note. Event-Related Potentials (ERPs) were time-locked to the voiceover onset of the relevant target location (graphically displayed as dotted line in the fourth photograph). At the end of the test trial, following the story scenario, participants verbally reported the answer to the question: “Where will Leo look for the car?”.

Participants watched 96 test trials (48 per condition) and 24 control trials (12 matching the agent preference with the location of the toy and 12 mismatching) divided into three blocks, for a total of 120 trials over all blocks. Trial presentation was pseudo-randomized within block. Each block consisted of 40 trials, counterbalanced per condition (true and false belief), stimuli presented (four possible objects), location of the object (left or right), preference of the agent towards the object location (left or right), and control questions (matching and mismatching) positioning.

The two ERP-evoking words (*Eimer*, bucket, and *Koffer*, case, in German in the original version) were chosen for having a similar length, phonetic stress, and syllabic structure, and were audio-processed to be normalized in length, to ensure maximal comparability between trials. EEG was recorded continuously and ERPs were time-locked to the onset of the two words for a 2000ms duration window and preceded by a 500ms silence interval in the voiceover, before the relevant object location was stated, which served as pre-stimulus ERP baseline.

3.3.3 Procedure

The participants sat comfortably in a dimly lit, sound attenuated chamber while videos were presented on a computer monitor. Stimuli were presented on a 19" screen (Dell Inc.) at a resolution of 1280x1024 and a refresh rate of 60Hz. The picture's size on screen was 25.5 cm in height and 38 cm in width, with the agent's on screen occupying ca. 2/3 of the picture on a dark grey frame background, completely visible during inter-trial breaks.

The experiment was conducted at an average viewing distance of 90 cm from the screen and the height of the display was individually adapted to be at the centre of the visual field of the participant.

At the beginning of the experiment, participants were asked to pay attention to the following videos and reply the question at the end of each trial. No further instructions were given. The trial presentation was paced by an experimenter, sitting in the EEG chamber with the participant and recording via mouse click each trial response, verbally given by the participant. The task lasted on average 60 minutes (20 minutes per block). Stimulus presentation was controlled using Presentation (Neurobehavioral Systems Inc., Albany CA) software.

After the experiment ended, participants were asked to fill out a post-test debriefing questionnaire with open questions about the comprehension of task instructions and whether any strategy was followed to solve the task. Questionnaires were qualitatively evaluated by the experimenter to detect possible bias in solving the task and eventually exclude invalid participants from the final sample.

3.3.4 EEG Recording and ERP Processing

Each participant wore an EasyCap EEG cap (Brain Products GmbH, Gilching), individually sized to their head circumference with a set of 64 + 2 active electrodes. Two extra electrodes, mounted on the cap, were used as reference (placed at position Cz) and ground electrode respectively (placed at position AFz). Electroencephalogram was recorded continuously from the scalp with a sampling rate set to 500 Hz, online low cut-off filter at 0.01 Hz and high cut-off filter at 100 Hz (resolution/unit 0.1 mV). Impedance was kept below 10 kΩ throughout the experiment.

Offline analyses were performed using Brain Vision Analyzer software (Brain Products GmbH, Gilching). All recordings were re-referenced against common average reference during offline pre-processing analysis. Continuous raw EEG data were digitally high pass-filtered at 0.1 Hz/-24 dB and low pass-filtered at 30 Hz/-12 dB before applying automatic ocular correction using the Gratton et al. (1983) algorithm for ocular movements detection. F9 and F10 channels were used to monitor horizontal eye movements, while Fp2 was used, in combination with the 64^{th} electrode from the EEG cap placed under the right eye, to record vertical eye movements and blinks.

EEG data were segmented into 1300ms epochs, including 1200 ms after stimulus onset and 100 ms as pre-stimulus baseline. Only valid trials, i.e. trials were the answer given was in agreement with the agent belief independently on the original object location, were included in the analysis. Trials in which verbal response time exceeded a window of around 2 seconds from the final question (exactly 12 seconds from the ERP trigger) were excluded.

Segmented data were scanned for artefacts in three steps. First, for each participant, channels where more than 10% of data was marked with artefacts (defined as channels with amplitudes exceeding 50 µV or with a running average activity exceeding 50 µV) were removed, while channels with artefacts affecting only in between 5 to 10% of the data underwent spherical spline interpolation. Second, segments in which blinks were detected (defined as peaks exceeding ±50 µV) were automatically eliminated. Third, segments with other artefacts that were previously undetected were excluded via manual visual inspection of all trials by the experimenter.

Artefact-free segments were averaged separately for each experimental condition with an inclusion criterion of a minimum of 25 valid segments per condition and participant. Following artefact rejection, the average number of usable segments was 38.7 (SD = 5.69) in the true belief (range: 29 - 48) and 37.9 (SD = 4.68) in the false belief condition (range: 29 - 46). The number of useable segments per participant did not differ significantly between conditions, $t(29) = 1.18$, $p = .248$, $d = 0.21$. Artefact-free segments were baseline corrected, averaged per condition to derive ERPs, and exported as mean area amplitude per electrode within specific 50-ms time segments, as described below.

3.3.5 Data Analysis

Repeated measures analyses were carried out for each mean amplitude value of 50-ms time segments from the 250 to 1200ms post-stimulus time interval using SPSS 24.0 (IBM Corp. Released 2016). This time window was selected as the one most relevant for the two ERP components we were interested in based on previous Theory of Mind literature (Liu et al., 2009a; Meinhardt et al., 2011; Sabbagh & Taylor, 2000). Two series of ANOVAs were performed in order to examine exhaustively columns of scalp electrodes along the anterior-posterior and left-right axes of the head, following a grid analysis procedure previously applied in the false belief study by Liu et al. (2009b). A first grid analysis covered the midline electrodes (FCz, Cz, Pz), fronto-central electrodes (FC3, FC4), central electrodes (C3, C4) and parietal electrodes (P3, P4). This analysis encompassed electrodes traditionally included in false belief ERP tasks, to ensure comparability (Geangu et al., 2013; Liu et al., 2009b; Meinhardt et al., 2011; Zhang et al., 2013). A second grid analysis covered the midline electrodes (Fz, CPz, Pz), lateral frontal electrodes (F5, F6), lateral centro-parietal (CP5, CP6) and lateral parietal electrodes (P5, P6). This laterally focused analysis was included after visual inspection of the ERPs grand average, on the assumption that lateralized effects due to the specific language-based presentation of our stimuli could be present.

The data for each time bin were analysed with condition (true belief versus false belief), hemisphere (left versus central versus right), and caudality (anterior versus central versus posterior) as within-subjects factors. Significant results ($p < .05$) are reported only when the factor condition is involved, and were followed by post-hoc tests if appropriate. To correct for the risk of capitalization on chance (i.e., correcting for the number of tests of significance being performed) a minimum of two consecutive significant intervals ($\alpha < .05$) of 50 ms were accepted as significant. As 19 mean amplitude values were extracted, a total number of 19 x .05 = 0.95 intervals were expected to show a significant effect by chance. Using at least two consecutive significant time segments for the analysis ensured the number of significant intervals by chance alone to be reduced to .045 (18 x .05 x .05, for two intervals), i.e. below the significant criterion of .05 (for this procedure see Gomarus, 2010; Müller et al., 2016). The largest p-value, the smallest F-value, and the smallest partial eta-squared for effect size are reported when significant effects in successive time segments were found. When Mauchly's test for violations of sphericity was significant, we applied Greenhouse-Geisser correction and reported corrected p-values and degrees of freedom for any interested effect.

3.4 Results

3.4.1 Behavioural Results

Response accuracy to true and false belief trials was measured on a binary scale (0 = incorrect, 1 = correct) for a maximum of 48 correct answers per condition possible. Overall, all participants showed mid-high to high accuracy in performance (minimum 38, maximum 48 correct answers).

Paired t-test revealed the false belief condition (M = 45.17; SD = 2.18; range = 42 - 48) to be more difficult than the true belief condition (M = 46.43; SD = 1.19; range = 38 - 48) for our participants ($t(29) = 3.05$, $p = .005$, $d = 0.56$). In a similar fashion, the matching condition of the control questions revealed to be easier (M = 11.73 ; SD = 0.52) than the mismatching one (M = 11.3; SD = 0.88), probably due to the unexpected discrepancy between the agent preferred location and the actual location of the object ($t(29) = 2.44$, $p = .021$, $d = 0.45$).

3.4.2 ERP Results

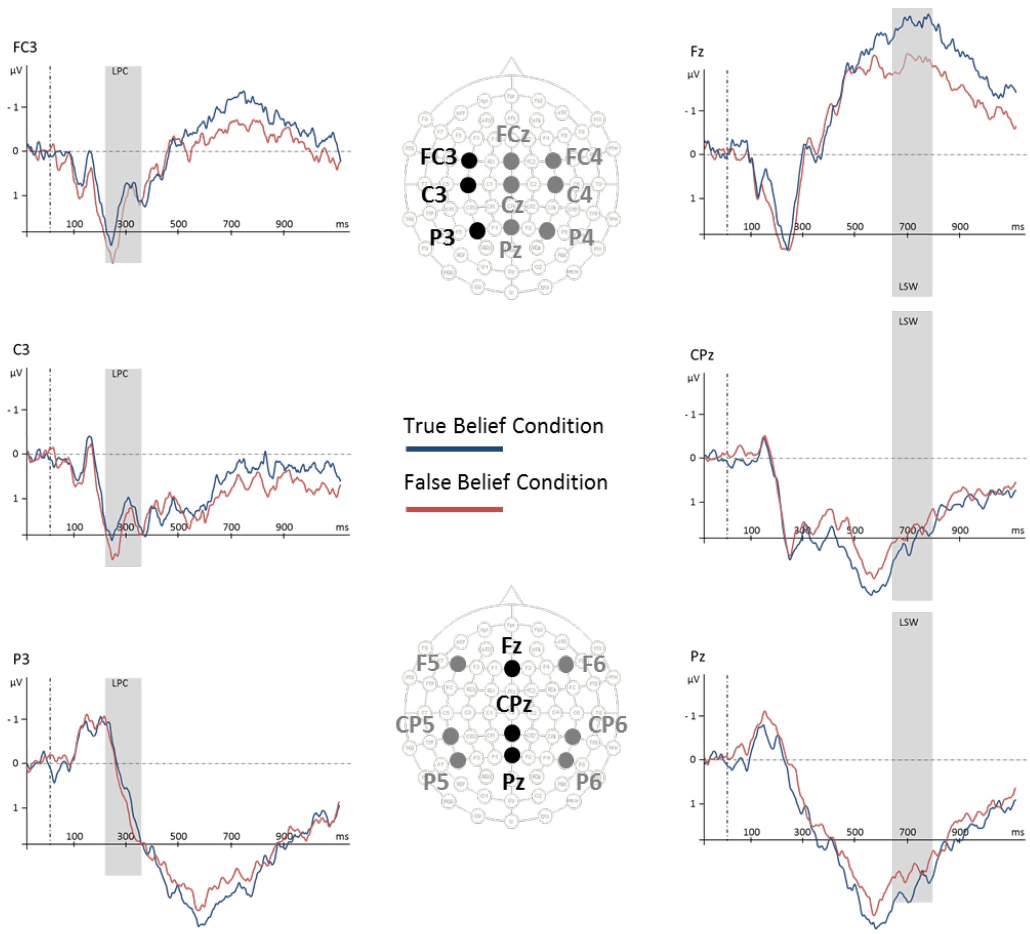
Visual inspection of grand averages waveforms (depicted in Figure 3.2 for representative electrodes) revealed that both true and false belief conditions elicited a sequence of early and late waveforms with a different distribution. A differentiation between the true and false belief condition was predominantly visible in two waveforms: the first deflection was most pronounced in the left hemisphere over midline and parietal region, while the second was bilaterally distributed at anterior regions (a whole scalp heat-map of the two time windows that showed significant results is showed in Figure 3.3).

A repeated measures ANOVA with condition, hemisphere, and caudality as within subject factors (following Liu et al., 2009b) revealed a significant interaction between condition and hemisphere between 250 ms and 400 ms after stimulus onset in our first grid analysis, encompassing three 50-ms time windows (F's(1.65, 47.79) > 3.37, p's $<$ .041, η_p^2's $> .104$). In order to shed further light onto the nature of this interaction, post-hoc pairwise comparisons of estimated marginal means were performed. Fisher's LSD test indicated that false belief condition elicited a more positive waveform than true belief condition in the left hemisphere, with a peak at 250 to 300 ms ($p = .025$, see Table 3.1 for descriptive statistics). This left-focused effect rapidly decreased between 300 to 350 ms ($p = .078$) and was shifted at inverse polarity to the centre of the scalp in the last significant time window, between 350 to 400 ms ($p = .042$). This pattern of results suggests that in the 250 to 400 ms epoch there was a lateralized component in the left hemisphere, characterized by a rapid shift from a more positive to a more negative waveform during the early stages of processing of false belief reasoning.

A more distinct waveform was identifiable in the second analysis we performed. A repeated measures ANOVA with condition, hemisphere, and caudality as within subject factors revealed a significant interaction between condition and caudality (anterior,

Figure 3.2

Grand average ERP waveforms based on grid analyses



Note. Grand average ERP waveforms at representative electrodes: frontal midline (Fz), fronto-central left (FC3), central left (C3), centro-parietal midline (CPz), and parietal midline (Pz) and left (P3).
A differentiation between false belief and true belief condition can be localized in the left hemisphere during the 250-400ms time window. A further differentiation between the two conditions was observable at midline electrodes in the 650-800 ms time window, with reverse polarity between anterior and centro-parietal regions.
Negative amplitude plotted upwards.

Table 3.1

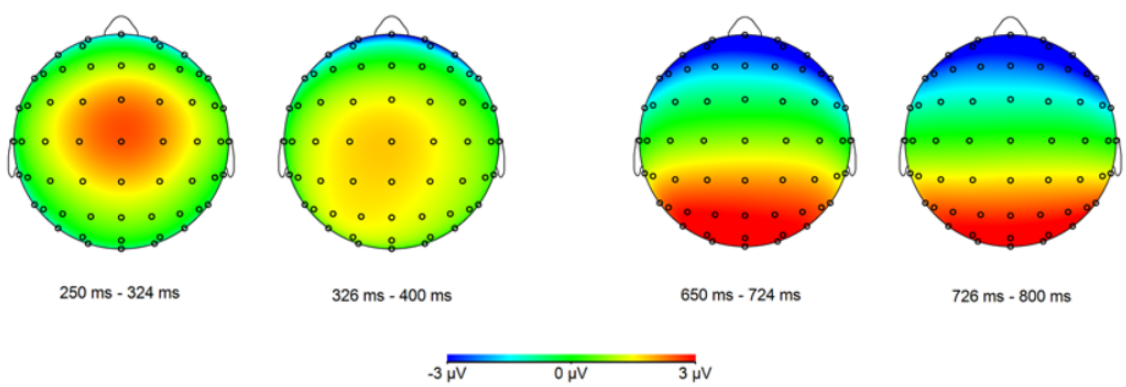
Summary of Mean Area Amplitudes at 250-350ms

Time Window	Left Hemisphere		Laterality Midline		Right Hemisphere	
	TB	FB	TB	FB	TB	FB
250 – 300 ms	0.89 (0.18)	1.27 (0.20)	2.21 (0.22)	2.03 (0.23)	1.09 (0.16)	0.97 (0.18)
300 – 350 ms	0.78 (0.22)	1.02 (0.20)	1.54 (0.22)	1.32 (0.24)	0.74 (0.17)	0.88 (0.18)
350 – 400 ms	1.33 (0.23)	1.31 (0.24)	1.71 (0.22)	1.24 (0.27)	0.89 (0.17)	0.80 (0.19)

Note. Mean Area Amplitude (in µV) over all participants (n = 30) in the subsequent 50-ms time windows that showed a significant cross-interaction between hemisphere laterality (Left, Mid, and Right) and condition with the two levels true belief (TB) and false belief (FB).
Standard error presented in brackets.

Figure 3.3

Topographic Heatmap of Scalp Electrical Activity



Note. Topographic voltage maps of scalp electrical activity based on amplitude difference between false and true belief condition in the two significant time windows.
A diffuse and progressively left localized positivity can be seen during the 250-400ms time window (left). In the 650-800 ms time window (right), a differentiation between the two conditions is observable with reverse polarity between anterior (more negative) and centro-parietal (more positive) regions.

central, posterior) between 650 ms and 800 ms after stimulus onset, encompassing three 50-ms time windows (F's(1.07, 26.69) > 4.87, p's < .034, η_p^2's > .163). Post-hoc pairwise comparisons (Fisher's LSD test) indicated that at 650 to 700ms the false belief condition elicited a more positive waveform than true belief condition in the anterior regions (p = .042, see Table 3.2 for descriptive statistics).

In the same time window, the opposite polarity, with a true belief wave more positive than a false belief one, was found in the central scalp region (p = .023) and with marginal evidence in posterior regions (p = .057). The effect continued in the 700 to 750ms time window, where false belief condition consistently elicited a more positive wave than true belief at anterior sites (p = .021), while opposite polarity was visible at central (p = .032) and posterior sites (p = .023). The same behaviour held significant for anterior (p = .040) and central regions (p = .022) in the last time window between 750 and 800ms, with a trend to significance for the posterior regions (p = .051). Due to its positive waveform and frontal localization, with reverse polarity on more posterior areas, this component showed similarities to the LSW wave found in previous ERP studies.

Table 3.2

Summary of Mean Area Amplitudes at 650-800ms

	Caudality					
Time Window	Anterior		Central		Posterior	
	TB	FB	TB	FB	TB	FB
650 – 700 ms	-2.60 (0.38)	-1.85 (0.38)	1.75 (0.20)	1.30 (0.19)	2.68 (0.32)	2.09 (0.29)
700 – 750 ms	-2.69 (0.39)	-1.89 (0.40)	1.67 (0.22)	1.26 (0.20)	2.63 (0.33)	2.02 (0.30)
750 – 800 ms	-2.81 (0.37)	-2.06 (0.36)	1.64 (0.21)	1.16 (0.19)	2.63 (0.31)	2.06 (0.26)

Note. Mean Area Amplitude (in µV) over all participants (n = 30) in the subsequent 50-ms time windows that showed a significant cross-interaction between hemisphere caudality (Anterior, Central, Posterior) and condition with the two levels true belief (TB) and false belief (FB).
Standard error presented in brackets.

3.5 Discussion

In the present study, we investigated the neural correlates of false belief understanding in a reduced processing demands task previously used only in behavioural research (Wellman et al., 2001). Aim of the study was to address specifically the cognitive processes correlated with mental states reasoning and test for generalizability of ERP findings across false belief paradigms. In order to achieve this goal, we developed an ERP-adapted version of a facilitated false belief paradigm, where the knowledge of the world from the agent's perspective is explicitly stated, reducing the effort to infer the representation of the agent's mental states in contrast with the real state of affair. In addition, the distinction between true and false belief was given uniquely by verbal information about the agent's knowledge of the world, reducing extraneous cognitive load often associated with the change of location of traditional unexpected transfer task.

At the behavioural level, our adult sample showed overall high accuracy rates in solving true and false belief trials; the same held true for the matching and mismatching preference control questions. Written post-test debriefing indicated that participants did not follow any specific pattern or strategy to solve the task, besides paying attention to the agent's belief in contrast to the real object location. The participants' focus on the representation of the agent's mental states for consciously solving the task corroborates our paradigm suitability to elicit false belief reasoning.

At the neurophysiological level, our findings show that two waveforms distinguished false and true belief reasoning in the trials' critical moment at distinct latencies and scalp distribution. We found evidence of an early positive wave at similar latencies of the LPC, previously identified in ERP paradigms focusing on false belief reasoning (Cao et al., 2012; Meinhardt et al., 2011; Zhang et al., 2009). The latency of this component supports its interpretation as a P3a-like wave evoked by managing mismatching information between the real location of the object and the false belief held by the agent and requiring inhibitory control over the participant's own knowledge of reality (Zhang et al., 2009). Although the spatial resolution of EEG prevents an accurate localization of neural events, the LPC lateralization in the left hemisphere could be due to the fact that our task relies heavily on language processing, in particular on auditory perception and comprehension of verbal stimuli. The ERP literature shows that language-related components are often lateralized (Friederici, 2002; Kutas & Federmeier, 2011), mirroring the results obtained with neuroimaging techniques in support of the role of the left prefrontal cortex as a hub for integration of linguistic meaning and one's own knowledge of the world (Hagoort et al., 2004).

Our reduced processing demands task evoked a second component at later latencies, characterized by opposite polarity between anterior and more central-posterior regions. Because of its distinctive latency and positive deflection in frontal regions during false belief processing, we identified it as a LSW, consistently reported in previous false belief studies independently of the presentation mode (Geangu et al., 2012; Meinhardt et al., 2011; Sabbagh & Taylor, 2000). This is consistent with the interpretation of the LSW as a neural marker specific of false belief reasoning, associated with decoupling the false agent's representation of the world from one's own true representation, in order to correctly solve the task (Liu et al., 2004).

Altogether, our results show that neural markers are identifiable at distinct latencies, possibly reflecting subsequent stages of cognitive processing of false belief understanding (Meinhardt et al., 2011). The presence of two waveforms, the first presumably connected with automatic processing of mismatch detection in the incoming information stream (as reflected in a P3a-like component, see Polich, 2007) and the latter elicited by controlled mental operations on internal mental states' representation, seems to define the temporal timeline of neural Theory of Mind processing. Given the similarities with previous findings from different paradigms, such as the unexpected transfer task, the pattern of neural activation emerging from our study design does not appear to depend on story format (i.e., multi-steps photographic scenarios relying specially on verbal comprehension). The processing of the sole discrepancy between false belief versus the state of reality seems instead to be at the basis of this pattern of neural activation, thus supporting the view that the reduced processing demands task, where information about the agent's mental states is overtly available, also requires and elicits false belief reasoning.

This is in agreement with findings of comparable difficulty from behavioural studies in children, where it has been found that low reduced processing demands tasks where belief is explicitly stated and standard false belief tasks are indeed correlated in performance results (Wellman et al., 2001).

In addition, our findings add to previous studies about the neural basis of false belief comprehension and support the view that reduced processing demands tasks, without the need to infer false belief from contextual information, evoke consistent components compatible with LPC and LSW, which define the temporal processing of false belief understanding and are in agreement with traditional false belief studies. Thus, the detected effects seem to be robust, as they showed to be insensitive to superficial task variations.

Nevertheless, as the literature on the neural underpinnings of false belief understanding is still scarce, more research must be done to investigate the generalizability of these ERP findings across other paradigms with various processing demands used in behavioural research. In particular, ERP literature on false belief is not exempt from contrasting interpretations regarding the specific cognitive processes associated to these neural markers, a struggle which points to the complexity of dissociating the LPC and the LSW. An interesting approach comes from a language comprehension study by Roehm et al. (2004), which showed that non-distinguishable slow ERP components can be disentangled in activity correlating to linguistic problem detection (theta band) and conflict resolution (delta band) by frequency-based decomposition. Future research could focus on employing this paradigm to investigate slow frequency oscillatory characteristics of false belief discrimination.

Lastly, given the suitability of the reduced processing demands task for testing false belief understanding in a short-story format, lower in working memory requirements and cognitive load, we plan in the future to extend this investigation to the children population, in order to explore the neural correlates of false belief at a crucial time in development for the emergence of Theory of Mind abilities.

Chapter 4

Study C: 27-month-olds' Early Sensitivity to Mental States Language Comprehension: An Eye-tracking Study

4.1 Abstract

The third year of life is a transitional period from an implicit to an explicit Theory of Mind. Research suggests a role of mental states language in fostering this conceptual passage during infancy. While there is initial evidence for the production of epistemic terms in the third year of life, the comprehension of *know* and *think* has not been systematically tested at this early age yet.

In the present study, we developed an eye-tracking task to investigate the implicit differentiation of these two epistemic verbs in 27-month-old children. Children (N = 146) participated in a hiding game in which two agents expressed contrasting epistemic statements about the location of a hidden object ("I know it is in there" versus "I think it is in there").

Fixations at the agents and the object's locations in the two seconds interval following the critical mental terms and after a final test question were recorded across eight trials. Children's implicit understanding of mental states verbs was measured via preferential looking time to the target (the agent that knows) compared to the distractor (the agent that thinks).

Growth curve models of the time course of target recognition across three test phases on preferential looking time analyses revealed that children differentiated between *know* and *think*. Unexpectedly, they had a consistent preference towards the agent uttering the epistemic state *think*.

We discuss potential explanations of this result in terms of a novelty effect in language acquisition processes. Our findings indicate a beginning sensitivity for implicit understanding of degrees of certainty in epistemic language in infancy.

Keywords: Theory of Mind, Preferential Looking, Epistemic verbs, Eye-tracking

4.2 Introduction

The acquisition of mental states terms in infancy has long been a matter of investigation in both the realm of linguistics and developmental psychology. Acquiring epistemic language (i.e., language expressing a speaker's evaluation of the likelihood of a state of affairs, see Yalcin, 2007) has been connected to a first understanding of conceptual representation of the mental states of others, as well as self-awareness of one's own ignorance. In particular, mental states verbs have been under scrutiny for their meta-representational features: in order to understand and use them correctly, children must first create a mental representation of the attitude involved (such as one's status of knowledge) and the content of the represented state (that is, what is known) in someone else's mental perspective (Antonietti et al., 2006). It follows that mastery of mental states language and meta-representation of beliefs has been considered an important step for developing predictions of knowledge (attributing truth propositions and acknowledging sources of knowledge) necessary for a transition to a full-fledged Theory of Mind, defined as the ability to recognize others' desires, goals, and perspectives (Montgomery, 1992; San Juan & Astington, 2012).

Up to date, research on epistemic verbs mainly focused on two main approaches: observational studies assessing the frequency of mental states verb usage in young children (focusing on explicit knowledge) and experimental studies on the comprehension of semantic differences in mental terms (implicit knowledge). Evidence from the first approach supports the view that the acquisition of mental states terms proves to be more difficult than other verbal categories, due perhaps to the representational ability necessary in evaluating the probabilistic nature of verbs like *know* or *think* (Johnson, 1982). In particular, the epistemic modality follows a developmental progression from two and a half to seven years of age, until children finally understand certainty expressions such as *think* and *believe* (Bartsch & Wellman, 1995; Bassano, 1996). In apparent contrast with these findings, experimental research indicates that children at a younger age already show non-verbal sensitivity to epistemic states in communication, such as 12-months-olds understanding of the knowledge state of others by providing informative pointing gestures to help ignorant but not knowledgeable partners (Liszkowski et al., 2008). This and other studies on early understanding of others' goals, intentions, and beliefs (e.g., Onishi & Baillargeon, 2005; Surian et al., 2007; Tomasello & Haberl, 2003) suggest that sensitivity to others' mental states emerges earlier than previously thought.

This recent evidence on an early sensitivity to mental states, together with the finding of spontaneous production of epistemic factive versus non-factive verbs in the third year of life (Harris et al., 2017) raises the question whether also mental states language comprehension can already be found in infancy. As epistemic terms exhibit an intrinsic pragmatic nature rooted in the social world we live in, an approach that combines the purely semantic aspects of mental states verbs with their functional context of use should be favoured in the attempt to investigate how and when the conceptualization of others' mental states comes into being during development (Matthews, 2014; Nuyts, 2001).

In the realm of language acquisition, several well-established paradigms have been employed to investigate implicit comprehension (Copeland & Gedeon, 2013; Conklin & Pellicer-Sánchez, 2016; Koornneef & Van Berkum, 2006), but mental states language understanding is under-researched. One reason for this is probably the difficulty to implement the situational cues necessary to track the understanding of mental states within context in an experimental setting. For instance, a classic word-learning paradigm is not informative about the semantic of a non-factive verb like *think* or *know*. Thus, in order to successfully investigate mental states understanding, the researcher must deploy a paradigm in which a situational setup is presented to the children so that they can rely on linguistic cues to solve the task.

To reach this goal, we developed a novel eye-tracking task, based on an established epistemic discrimination task for 3.5-year-old children (Moore et al., 1989; Kristen-Antonow et al., 2019). To test younger children, we measured the discrimination of *know* versus *think* via preferential looking (see Skoruppa et al., 2013), a well-established procedure in eye-tracking research in infancy to assess cognitive states during language processing (Blom & Unsworth, 2010).

Thus, we developed a procedure in which two agents expressed contrasting epistemic statements about the location of a hidden object, consisting of an initial behavioural familiarization and followed by the eye-tracking version of the hiding game.

During the familiarization phase, the child learned to recognize and evaluate contrasting, non-epistemic knowledge statements of two agents about the location of a hidden object (based on the Moore et al., 1989 paradigm). The child was encouraged to listen carefully to the statements of both agents before helping the experimenter retrieving the object from the correct of two possible locations. During the eye-tracking part of the task, the same instructions were given to the child but the agents' statements were now epistemic in nature (*know* versus *think*). The video sequence was separated in three test phases, corresponding to each of the statements uttered by the two agents on screen regarding their opinion about the hidden location of the object, as well as to a final test phase in which a question regarding the actual location of the object was directed to the child by an out-of-field voice.

Infants at 12 months of age already have understanding of others as intentional and attentional agents (Tomasello & Haberl, 2003) and, by means of preferential looking paradigm, it has been shown that at this early age they are capable to simultaneously look in anticipation of action-oriented goals (Carpenter et al., 1998). Hence, in our hiding game, we expect the relative looking duration spent towards the two agents in the act of expressing their opinion on the object true location to be implicit indicator of attention towards the agents' pointing behaviour.

In particular, if non-verbal sensitivity to epistemic states emerges already in early childhood, children should show a preferential looking pattern towards the agent holding the highest degree of certainty (*know*) about the object location during the second test phase, when both statements are available to the child's mental states processing. A similar effect could be already visible in the first test phase, in case anticipation mechanisms due to the repeated nature of the task come into place (i.e. child consistently looking with anticipation to the agent holding the highest degree of certainty, as soon as the first opinion is stated, independently on the semantic nature of the statement itself).

Finally, eye-tracking studies show that infants undergo substantial development in visual short term memory capacity for locations already in the first year of life (Oakes at al., 2011; Oakes at al., 2013) and exhibit consistent anticipatory looking in recalling from long term memory from 18 months of age on (Nakano & Kitazawa, 2017). Thus, if children at 27 months of age have already an implicit understanding for epistemic states, we expect that they will withhold information connected to the most trustworthy agent until the end of the trial, and exhibit a consistent preferential looking pattern towards the location associated to the object true location also in the final test phase.

4.3 Methods

4.3.1 Participants

Participants in this study were 146 infants with a mean age of 27 months and 6.87 days (SD = 9.46 days; range = 26 months and 6 days – 28 months and 6 days; 77 female and 69 males). Twenty additional infants were tested but excluded due to fussiness (n = 3), bilingualism (n = 1), technical difficulties (n = 1), or insufficient language skills as measured by our language battery (n = 15). Two subsamples were generated for independent test-phase analyses by removing participants that did not provide at least one valid trial per condition (salient agent knows or thinks) in T2 (n = 10) and at least two valid trials overall in T3 (n = 9). This choice was made as the last test phase was the most relevant for our study, being the only one not biased by the saliency of the last speaking agent.

No participant had less than one valid trial per condition in T1. None of the infants had any known hearing or visual problems, and all came from homes where German was the language mostly used in their daily routine (at least 70% of the time). Children's semantic and syntactic comprehension was measured with a German standardized test for language development at two years of age (SETK2, see Grimm, 2000). Information on the children's mental states language production was assessed with a parent-administered mental states language questionnaire (MSLQ, see Bretherton & Beeghly, 1982).

The caregiver gave informed written consent before starting the experiment and received monetary compensation for travelling expenses. Children received a gift for their participation.

This study was approved by the local ethics committee. All procedures performed in the reported experiments were in accordance with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

4.3.2 Experimental Paradigm

For the familiarization phase of the hiding game, we used two coloured square boxes (9x9x3 cm, blue or green), a white hiding screen (60x50 cm), and two monkey puppets (height: 12 cm) each with a different shade of fur. Eight animal stickers (cat, dog, cow, sheep, duck, rabbit, pig, and chicken) were selected as objects to be hidden in one of the boxes during the game.

For the experimental phase of the hiding game, we adapted the paradigm used by Moore et al. (1989) to an eye-tracking setting. We created approximately 35-second-long cartoon movies animated by vector graphics elements on a grey neutral background (see Figure 4.1 for a trial example including the time windows under analysis). Prior to the first trial, an introduction video was shown in which the narrator stated "Look at the monkeys! They will help you to find the sticker." At the beginning of each trial, a narrator (not pictured on screen) presented a sticker in her hand and expressed her intention ("I will hide this sticker!"). The animal stickers in the videos were the same the child played with during the behavioural familiarization phase. All voiceovers used in the videos were recorded in German by female experimenters in child-directed speech.

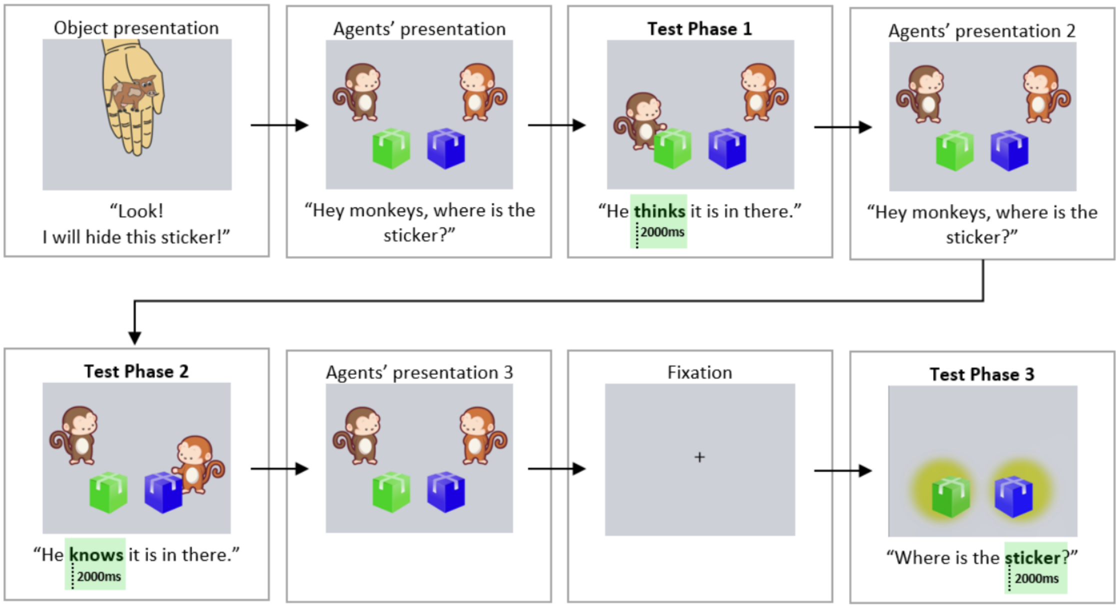
In the next scene, two monkeys standing next to one of the two boxes respectively were presented. The narrator asked: "Hey monkey, where is the sticker?". Following this question, one of the two monkeys started jiggling and moved towards the box next to it and replied: "Here!". After the first monkey spoke, a second narrator was introduced and gave a statement about the knowledge of the monkey, expressing a high ("He knows that it is in there") or low ("He thinks that it is in there") degree of certainty. After the first monkey came back to its original position, the first narrator asked again about the hidden object's location. Now, the second monkey jiggled and moved towards the box next to it. After the second monkey's reply ("Here!"), the second narrator gave a statement about the degree of certainty of this monkey, always opposite to the first monkey's statement.

A screen with the same grey background and a central fixation cross was inserted to re-focus the attention of the child to the screen before the final part of the video. The screen was paced by the experimenter by mouse-click, letting the video continue once the child focused on the fixation cross. In the final sequence, a steady image of the two boxes was presented. The boxes were highlighted by two bright light circles flashing for two seconds during the final prompt question: "Where is the sticker?". The monkeys were removed from the final scene to avoid biases in gaze behaviour of the child due to the expectation, found in a previous pilot phase of the experiment, that one of the monkeys will speak again.

The colour of the boxes, the order of statements referred to the monkeys degree of knowledge, as well as the order of which monkey was speaking first was counterbalanced across trials.

Figure 4.1

Mental States Language Comprehension Task



Note. Trial example from the epistemic verbs comprehension eye-tracking task. Goal of the task was to find the hidden object with the help of two agents. After initial presentation of the object that was going to be hidden, the two agents were presented. Three relevant test-phases were under investigation: T1 and T2 were time-locked at the onset of the relevant verb expressed by the agent speaking in that test phase (thinks/knows, counterbalanced across trials). After an experimenter-paced fixation screen, the final test phase T3 was time-locked at the onset of the salient word indicating the hidden object (sticker).

4.3.3 Procedure

Familiarization Phase

At the beginning of the session, the child sat alone or on the caregiver's lap at a table. The first experimenter sat in front of the child, while a second experiment stood at the side. The experimenter instructed the child that they would play a hiding game, in which the child had to find a hidden sticker in one of two boxes. In a first introductory trial, the experimenter showed a sticker to the child and hid it in plain sight.

This introductory trial was used to assess the child's capacity to follow the game's instructions and find a hidden object. Once the sticker was hidden, the experimenter asked the child to point to the box in which the sticker was located. In case the child did not point or chose the wrong box, the trial was repeated until the child correctly answered the question.

Following the introductory trial, four practice trials were presented. The first experimenter explained to the child that she was going to hide a sticker in one of the two boxes behind a visual barrier. Children were told that the monkeys would help them find the sticker and that they had to listen carefully to what the monkeys say, because this information would help them find the sticker. After the sticker was hidden and the visual barrier removed, the first experimenter addressed the monkeys asking them where the sticker was. The two monkeys, controlled by the second experimenter, moved forwards in turn and made a statement (uttering "Here!") indicating a box each. After each monkey's statement, the first experimenter commented either saying "It is really there!" in case of the target box or "He just pretends that it is inside" in the case of the empty distractor box. At this point, the experimenter asked the child to look for the sticker ("Where is the sticker?"). After the child's response, the selected box was open. If the chosen box was the incorrect one, feedback was given to the child by retrieving the sticker from the correct box. Every sticker used in the familiarization was given as gift to the child, independently of success.

Children were presented with four practice trials selected from a pool of 16 and counterbalanced across subjects for colour of the boxes, monkey speaking first, hidden location of the sticker, and statement associated with each monkey. The aim of the practice session was to familiarize the child with the goal of the task and the importance of paying attention to the monkeys' statements. From the practice sessions children learnt that the monkeys were not always right, a point of fundamental importance for being able to assess uncertainty and discriminate between the epistemic verbs used in the test phase.

Eye-Tracking Phase

After the familiarization phase, the participants moved to the eye-tracking setup, where they sat comfortably in a child safety seat or on their caregiver's lap. The experiment was conducted at an average viewing distance of 60 cm from the screen, based on eye-tracking calibration preceding the experimental recording.

Gaze data were recorded with a Tobii T60 eye tracker (Tobii Technology, Sweden) at 60 Hz. The stimuli were presented on a 17" display integrated into the eye tracker with a 1280x1024 video screen resolution. Both stimulus presentation and data acquisition were performed using the Tobii Studio software (Tobii Technology, Sweden).

The experimenter told the child that they were going to watch some animated movies where two monkeys would help them to find hidden stickers. No further instructions were given.

Children watched two blocks (around 10 minutes in duration), consisting of intermixed fillers and test trials, for a total of eight test and four filler trials. In filler trials, the second narrator gave statements similar to the ones used in the familiarization phase ("He just pretends that it is inside" or "It is really in there!") after each monkey moved towards the respective box. At the end of the filler trials, the box opened revealing the real location of the sticker, in agreement with the statement of the monkey who was not pretending. Filler trials were introduced to reward children via feedback (the appearance of the sticker), in order to retain their motivation to look for the sticker throughout the whole block, even in test trials where they did not receive feedback about the location of the sticker.

Order of presentation of trials was randomized and the ordering of blocks across participants was counter-balanced to avoid possible order effects. A break was introduced in-between the two blocks to ensure a resting pause for the child before the second part of the recording began. Calibration was performed a second time prior to the onset of the second block. A total of four attention grabbers were inserted across trials to prompt attention to the experimental videos.

4.3.4 Data Pre-processing

In the familiarization phase, the child's pointing or verbal response when asked where the sticker was hidden was transcribed online by a second experimenter. An independent coder confirmed the accuracy of online coding by rescoring children's performance via video recordings of the session. Each child was given a score ranging from 0 (no trials correct) to 4 (all trials correct). A score of 2 indicated performance at the chance level.

In the eye-tracking phase, gaze behaviour was recorded and segmented into three time windows during each trial. Fixations were identified using a velocity-based filter (Salvucci & Goldberg, 2000). A fixation was defined as all consecutive gaze samples with a velocity of about 30 deg/s or less and at least 80 ms in duration, with a sampling data rate of 16.7 ms.

The first two test phases (named T1 and T2 in Figure 4.1) corresponded to 2 seconds time windows following the onset of the mental states verb (i.e. *knows*, defined as target; or *thinks*, defined as distractor) used by the narrator to refer to each monkey's degree of knowledge. Thus, T1 began at the onset of the mental verb following the first monkey speaking, and T2 began at the onset of the mental verb following the second monkey speaking in any given trial. The third test phase under investigation (named T3) corresponded to a 2 seconds time window following the onset of the salient word "sticker" in the final question "Where is the sticker?". The timing for the third time window was based on the observation from a pilot phase that task-relevant gaze behaviour was triggered at the onset of the salient word “sticker”, when the child starts performing visual search towards the two boxes.

An initial baseline for each test phase was marked in a 2 seconds time window before the onset of each mental states verb, but it was not included in subsequent analyses as it did not bring any useful information to evaluate the child’s gaze preference prior to the onset of the test phase. It was in fact observed that the attention-grabber bell sound, the jiggling, and the following movement towards the box of the monkey before it started to utter its statement attracted the child attention, biasing the neutrality of the baseline phase towards the last speaking agent. Instead of making use of the baseline information, we therefore recoded children gaze behaviour discriminating towards the last speaking agent that *knows* or *thinks* in the exploratory part of our data analysis (see Data Analysis section).

Pre-processing of data (segmentation of raw recordings in test phases, creation of time-sequence scenes, and selection of areas of interest (AOIs)) was performed in Tobii Studio. Test phases in which the child did not pay attention to the sequence of statements were removed after visual inspection of video recordings by two independent coders. This led to a higher number of T1 (M = 6.33; SD = 1.59) compared to T2, and higher T2 (M = 5.54; SD = 1.8) compared to T3 (M = 4.98; SD = 1.84) usable trials, with a range from one to eight maximum trials available for analysis per test phase. The first 240 ms at the onset of the first fixation before the start of each test phase were removed from the analysis to compensate for the processing time needed for saccade initiation in young children, due to the uncertain nature of the first fixation and the impossibility to clearly link it with task-related behaviour (Swingley et al., 1999; Von Holzen & Mani, 2012).

In similar fashion, the last 40 ms of each test phase were removed from analyses to avoid skewed gaze patterns due to imperfect timing between the eye-tracking sampling rate and the segmentation of time windows. Relevant AOIs for the analyses were defined as the region around the two boxes and the monkeys' upper body and face. Four AOIs were defined: left and right monkeys in T1 and T2 (360 x 250 pixels, if monkey was moving closer to the box, 300 x 300 pixels if monkey was standing still), left and right boxes in T1, T2, and T3 (360 x 360 pixels). Proportions of AOIs were chosen to cover equal surface areas.

Gaze duration over AOIs in relevant time windows was then exported to R (version 3.6.2 (2019-12-12), R Core Team, 2020) for further data processing and statistical analyses. Separately for each phase, the proportion of infants' fixations on target (the monkey that *knows*) and distractor (the monkey that *thinks*) was calculated in order to get a proportion of target looking time (i.e. PTL = target / target + distractor, see Mani & Plunkett, 2007, 2008, for similar analyses). The resulting value ranges from 1 (looking only at the target) to 0 (looking only at the distractor) allowing us to assess whether participants fixated the target in preference to the distractor across trials.

Because of some children's preference for the monkeys (as most salient objects on the screen) and some towards the boxes (as objects of reference of monkeys' pointing behaviour), we merged AOIs of monkeys and boxes on the same side of the screen for T1 and T2. The combined AOIs of monkey + box (referred as "agent") conveyed the highest amount of valid PTL data. In T3 the PTL scores were given only by boxes AOIs, as no monkeys were present on screen.

4.3.5 Data Analysis

In order to test the validity of the experimental data, two analyses were performed. First, a one-sample t-test was used to assess the overall difference in the proportion of looking time to the target (the monkey that *knows* or the box that the monkey that *knows* pointed to) score against chance level across all test phases of the epistemic task in the predefined 2 seconds time window. This analysis was informative to exclude the possibility that the recorded PTL score could reflect just random variation in the children's gaze.

Secondly, after visual inspection of the children's overall gaze behaviour, we noticed a bias towards the agent that was speaking in the test phase under analysis, independently on whether the agent was the one that *knows* or *thinks*. The greater saliency of the agent that had most recently moved or spoken unequivocally attracted children attention in T1 and T2. In order to account for this bias, we reanalysed the data coding the agents based on whether the last speaking one was the one that *knows* (target) or the one that *thinks* (distractor) and we analysed the proportion of fixations to the target across this new saliency condition. Using a growth curve model we examined whether children looked more towards the agent who had just spoken when this agent knows where the sticker is compared to when this agent thinks the sticker is in there. The same rationale was used for T2 analysis.

In particular, we applied a generalized linear mixed model (GLMM, see Bates et al., 2015) to explore the gaze dynamics over time towards the most salient agent (speaking agent or box pointed at by that agent) during the first two test phases. We fitted the generalized mixed model on PTL scores using lme4's *lmer* function in R (Bates et al., 2015) with Gaussian error structure and identity link function (for a detailed description of growth curve model analysis see Mirman, 2015). An initial full model included time and condition (i.e. the speaking agent during the test phase under analysis) with the two levels *know*/*think* as fixed effects of interest, while participant id was included as random factor to allow for the estimation of a certain degree of randomness in the slopes across participants. Next, a reduced model was fit that did not include the factor condition.

ANOVA model fit comparison between the reduced and the full model allowed us to estimate the influence of the saliency of the speaking agent in T1 and T2. When the best fitting model is the full one, we can conclude that the factor condition has a unique effect on gaze behaviour and needs to be accounted for when investigating the impact of *know* and *think* on PTL scores. We also included time and its quadratic and cubic polynomial in the model, in order to investigate eventual non-linear changes in looking behaviour across the whole test phase duration. Successful preference for the target is assumed to reflect either in the linear or in the quadratic and cubic polynomials, as their inverse u-shaped curvature expresses the looking behaviour of the infant when looking to the target agent/box (i.e. first at chance level, then looking more to the target upon hearing the onset of the mental verb, then going back to chance level). If the children were able to discriminate between *know* and *think* independently on the agent that was speaking in that test phase, we expect the factor condition to interact with the respective time polynomials, which would reflect different curvatures for the *know*/*think* factor levels in T1 and T2. The factor condition was not included in the GLM for T3, as the experimenter-paced fixation screen and the following static scene prevented a saliency bias. Thus, we generated only a full model with time as fixed factor to explore gaze dynamics to the boxes associated with *know* and *think*. Data diagnostics for all the GLMMs was performed by visual inspection of QQ-plots and residuals' histograms, showing normal distribution of data points.

4.4 Results

4.4.1 Behavioural Results

In the familiarization phase, children correctly selected the box in which the sticker was hidden in 56.72% of the cases. On average, children passed two out of four trials (M = 2.26; SD = 1.11). Two children did not pass the first practice trial and were administered with additional repetitions before the familiarization phase started.

4.4.2 Eye-Tracking Results: Test Phase 1

A one-sample t-test was conducted to assess the proportion of fixations to the target, representing the proportion of time children spent looking at the agent that *knows* (the target) versus both agents (i.e. the agent who *knows* and the agent who *thinks*, the distractor). The proportion of fixations to the target was not significantly different from chance level set at 0.5, $t(145) = -0.20$, $p = .837$, $d = -.02$.

These results suggest that children did not have a specific preference towards target or distractor in T1 (see Table 4.1 for descriptive statistics for all test phases). Following our observation that the saliency of the speaking agent could hinder children preference (see "Data Analysis" in the Methods section), we proceeded with a GLMM exploratory analysis of the proportional looking time to the most salient agent.

The growth curve model comparison between a full model including the fixed factors time and condition, i.e. the speaking agent *knows* or *thinks*, and a reduced model including only time as fixed factor was significant, with the full model exhibiting a better fit to the data ($X^2 = 126.76$, $df = 4$, $p < .001$, see Table 4.2 for reduced and Table 4.3 for full models statistics comparison).

Table 4.1

Summary of Mean Proportional Looking Times

	Proportional Looking Times		
Test Phase	Mean	St.Dev.	Sample (N)
T1	0.497	0.119	146
T2	0.49	0.135	136
T3 (Final)	0.473	0.133	137

Note. Descriptive statistics of proportional target looking time (PTL) across the three test phases. PTL ranges between 0 (preference for distractor) to 1 (preference for target), a score of 0.5 indicates chance level.

Table 4.2

Reduced General Linear Mixed-Model for Test Phase 1

	Reduced Model T1				
Factors	Estimates	*SE*	*95% CI*	*t*	*p*
Intercept	0.666	0.011	0.644 – 0.688	59.213	< **0.001**
Time (poly1)	0.139	0.098	-0.053, 0.331	1.420	0.156
Time (poly2)	-0.145	0.064	-0.271, -0.019	-2.248	**0.025**
Time (poly3)	-0.043	0.013	-0.068, -0.017	-3.262	**0.001**
N	146_{id}				
Observations	34633				
Deviance	-5207.493				
AIC	-5185.486				
log-Likelihood	2603.747				

Note. Reduced GLMM testing the effect of Time including its linear, quadric, and cubic term on Proportional Looking Times in Test Phase 1 (T1).

Table 4.3

Full General Linear Mixed-Model for Test Phase 1

	Full Model T1				
Factors	Estimates	*SE*	*95% CI*	*t*	*p*
Intercept	0.679	0.011	0.657, 0.701	59.884	< **0.001**
Time (poly1)	0.124	0.099	-0.070, 0.317	1.249	0.212
Time (poly2)	-0.153	0.066	-0.282, -0.024	-2.327	**0.020**
Time (poly3)	-0.082	0.018	-0.118, -0.046	-4.446	< **0.001**
Condition (know)	-0.025	0.002	-0.030, -0.021	-10.761	<**0.001**
Condition * Time (poly1)	0.031	0.026	-0.021, 0.082	1.166	0.244
Condition * Time (poly2)	0.016	0.026	-0.035, 0.068	0.620	0.535
Condition * Time (poly3)	0.079	0.026	0.028, 0.130	3.018	**0.003**
N	146_{id}				
Observations	34633				
Deviance	-5334.253				
AIC	-5304.239				
log-Likelihood	2667.127				

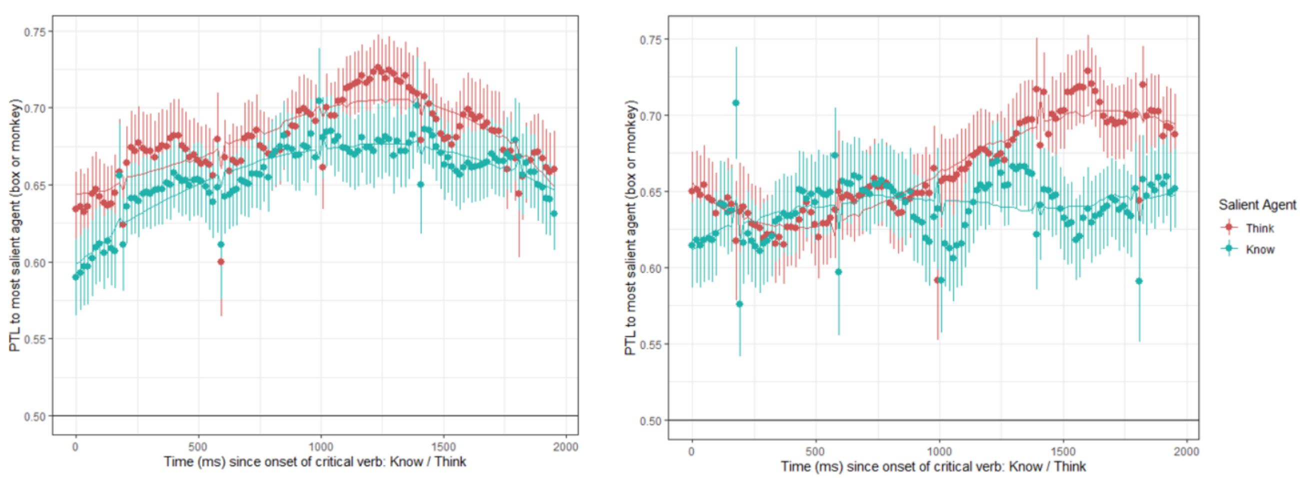
Note. Full GLMM testing the effects of Condition (salient agent *knows/thinks*) and its interaction with Time including its linear, quadric, and cubic term on PTL in T1.

Using the function *drop1* (Mirman, 2017), the full model revealed a significant interaction of condition with the cubic polynomial of time ($X^2 = 9.11$, $df = 1$, $p < .001$), reflecting differences in the time course of target recognition for *know* and *think*.

This suggests that participants showed a different pattern of looking behaviour when the agent that had most recently spoken knows where the sticker is, compared to when the agent only thinks that the sticker is in there. Looking at the curvature patterns of target fixations (graph on the left in Figure 4.2), we see that 27-month-olds had an overall higher preference towards the speaking agent when the speaking agent thinks compared to when the speaking agent knows where the sticker is.

Figure 4.2

Time course of PTL to Salient Agents in Test Phase 1 and Test Phase 2



Note. Participants' PTL to the most salient agent at the onset of the critical verb for T1 (left) and T2 (right). The lines reflect the fitted GLMM including time up to the cubic term for each condition, with the condition *know* expressed in blue and *think* in red. The line at 0.5 represents chance level, everything above it represents looking at the target.

4.4.3 Eye-Tracking Results: Test Phase 2

A one-sample t-test against chance level revealed that the proportion of fixations to the target in T2 was not significantly different from chance, $t(135)$ = -0.85, p = .399, d = -.07. These results suggest that children did not have an overall specific preference towards target or distractor in T2 (see Table 4.1). Since the saliency of the speaking agent potentially biased T1 as well as T2, we proceeded with a GLMM to investigate and control for the impact of agent saliency in discriminating *know* and *think*.

The growth curve model comparison between a full model including condition and a reduced model including only time revealed that adding condition as fixed effect significantly improved the model fit (X^2 = 206.27, df = 4, p < .001, see Table 4.4 for reduced and Table 4.5 for full models statistics comparison). In particular, the full model showed a significant interaction of condition with the linear (X^2 = 58.35, df = 1, p < .001), quadratic (X^2 = 10.50, df = 1, p = .001) and cubic polynomials of time (X^2 = 32.65, df = 1, p < .001). These results reflect differences in the time course of target recognition based on whether the salient agent in the second test phase was the one that knows or thinks. The graph on the right in Figure 4.2 shows two distinct gaze patterns due to the interaction of time and condition. At first, during the former 1 second part of the time window, children did not exhibited a specific preference for the salient agent that knows or thinks, but an increasing and consistent preference for the salient agent that thinks emerged later, during the latter 1 second part of the time window.

In order to shed further light onto this dynamics, we split the time window and performed an exploratory paired t-test between the amount of fixations to the speaking agent that knows and the one that thinks in the first and second 1-second-long parts of T2 separately. Paired t-tests revealed no differences between condition in the first part, $t(130)$ = -0.39, p = .699, d = .003, while a trend for a preference for the salient agent that thinks was found in the second part, $t(130)$ = -1.94, p = .054, d = .17.

Table 4.4

Reduced General Linear Mixed-Model for Test Phase 2

	Reduced Model T2				
Factors	Estimates	*SE*	*95% CI*	*t*	*p*
Intercept	0.649	0.014	0.623, 0.676	47.614	< **0.001**
Time (poly1)	0.177	0.105	-0.029, 0.383	1.681	0.093
Time (poly2)	0.009	0.075	-0.139, 0.157	0.122	0.903
Time (poly3)	-0.044	0.014	-0.071, -0.016	-3.091	**0.002**
N	136_{id}				
Observations	31741				
Deviance	-2515.344				
AIC	-2493.336				
log-Likelihood	1257.672				

Note. Reduced GLMM testing the effect of Time including its linear, quadric, and cubic term on Proportional Looking Times in Test Phase 2 (T2).

Table 4.5

Full General Linear Mixed-Model for Test Phase 2

	Full Model T2				
Factors	Estimates	*SE*	*95% CI*	*t*	*p*
Intercept	0.662	0.014	0.636, 0.689	48.257	< **0.001**
Time (poly1)	0.283	0.106	0.076, 0.490	2.680	**0.007**
Time (poly2)	0.055	0.077	-0.096, 0.205	0.711	0.477
Time (poly3)	-0.125	0.020	-0.164, -0.086	-6.270	< **0.001**
Condition (know)	-0.027	0.003	-0.032, -0.022	-10.470	< **0.001**
Condition * Time (poly1)	-0.218	0.029	-0.274, -0.162	-7.643	< **0.001**
Condition * Time (poly2)	-0.092	0.028	-0.148, -0.036	-3.241	**0.001**
Condition * Time (poly3)	0.161	0.028	0.106, 0.216	5.716	< **0.001**
N	136_{id}				
Observations	31741				
Deviance	-2721.617				
AIC	-2691.602				
log-Likelihood	1360.809				

Note. Full GLMM testing the effects of Condition (salient agent *knows*/*thinks*) and its interaction with Time including its linear, quadric, and cubic term on PTL in T2.

At the onset of the verb uttered in T2, children focused on both mental verbs, independently on whether the agent speaking was the one that knows (M = 0.63, SD = 0.25) or the one that thinks (M = 0.64, SD = 0.25). After this initial indiscriminate moment of attention for the speaking agent, children's focus prominently switched towards the agent that thinks (M = 0.69, SD = 0.24). When the agent speaking was the one that knows, children kept not having a specific preference (M = 0.64, SD = 0.23).

4.4.4 Eye-Tracking Results: Test Phase 3

A one-sample t-test to assess the proportion of fixations to the target versus distractor box revealed a significant difference from chance level (t (136) = -2.35, p = .020, d = .20), with children showing a consistent preference towards the distractor box (see Table 4.1). This result was confirmed by a growth curve model including time as fixed factor, which showed a trend to significance for a negative main effect of the linear polynomial of time ($X^2 = 2.76$, $df = 1$, $p = .096$, see Table 4.6 for model statistics). Visual inspection of the gaze pattern in T3 revealed a steadily increasing preference towards the box pointed at by the agent that thinks throughout the whole final test phase (see time course graph in Figure 4.3). This finding suggests that by the end of the trial, when the task-relevant final question "Where is the sticker?" was asked, children preferably fixated the box associated with the verb *think*, expressing a lower degree of certainty.

Table 4.6

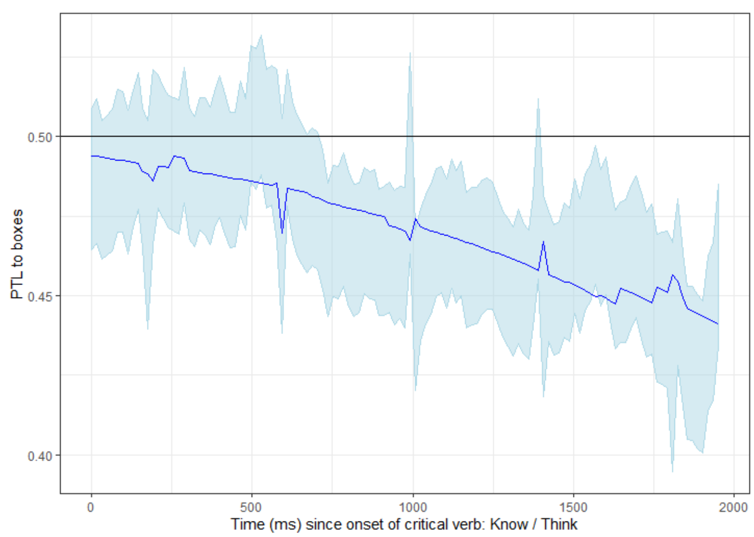
General Linear Mixed-Model for Test Phase 3

	Model T3				
Factors	Estimates	*SE*	*95% CI*	*t*	*p*
Intercept	0.469	0.011	0.448, 0.491	42.994	< **0.001**
Time (poly1)	-0.213	0.127	-0.463, 0.037	-1.670	0.095
Time (poly2)	-0.056	0.090	-0.231, 0.120	-0.623	0.533
Time (poly3)	-0.012	0.014	-0.039, 0.015	-0.835	0.404
N	137				
Observations	16556				
Deviance	-12126.926				
AIC	-12104.910				
log-Likelihood	6063.463				

Note. GLMM testing differences between boxes associated with know and think as defined by PTL including the effect of Time, its linear, quadric, and cubic term in T3.

Figure 4.3

Time course of PTL to Salient Locations in Test Phase 3



Note. Participants' PTL to the boxes associated with know and think during the final test phase. The blue line reflects the GLMM including time up to the cubic term. The line at 0.5 represents chance level, smaller values indicate preference for the distractor.

4.4.5 Correlations with Mental States Production

We collected data from the MSLQ questionnaire for a total of 157 children that took part in the experiment. Here we report the results from the Cognition category, which includes the production of mental states verbs such as *think*, *know*, and *believe*. In this age group, the average performance on the overall Cognition category was quite low (M = 22.2%; SD = 20). In particular, 32 children out of 157 were labelled by their parents as active users of the verb *glauben* (*think*, in German) while 111 were labelled by their parents as active users of the verb *wissen* (*know*, in German). When considering only the children that were included in the eye-tracking task (see Methods section for exclusion criteria), 30 children were able to produce *glauben* versus 101 who were able to produce *wissen*. To explore whether explicit production and implicit comprehension of mental states verbs correlate in this age group, we performed a point-biserial correlation between the PTL results from the final Test Phase 3 and the binomial output of the MSLQ. A significant negative correlation was indeed found between the score for the verb *wissen* (*know*) and children preference towards the target box in the third test phase of our eye-tracking task ($r(127) = -.21$, $p = .022$), showing that children that were labelled by their parents as already acquainted with the usage of *know* were also fixating more often to the distractor box. No significant correlation was found between PTL results and the verb *glauben* (*think*) score ($r(128) = .17$, $p = .861$), suggesting that productive competence (or lack thereof) for the verb *think* is not indicative of children implicit comprehension for the same verb at such early age.

4.5 Discussion

In this study, we investigated whether 27-month-old children already show implicit initial discrimination for mental states verbs. Our sample was presented with an eye-tracking version of a facilitated hidden object task in which they received contrasting statements from two agents, one that knew where the hidden object was located (the target of the task, indicating the highest degree of certainty) and one that only thought the hidden object was in a certain location (the distractor, indicating the lowest degree of certainty).

After pre-processing of the eye-tracking data, analysis of proportional looking time to the target versus distractor revealed no particular preference towards the target or the distractor in the first two time windows, i.e. when the first and the second agent (randomized during trials presentation) uttered their statements. Our interpretation for this finding is that children exhibited interest for both agents and were particularly attracted by the saliency of each moving and pointing agent per test phase. Thus, this led us into an exploratory analysis of the time course of target recognition, redefined as the salient (speaking) agent per time window. Once the bias towards the salient agent was taken into account, a consistent preference for the agent that thinks emerged from the analysis of the first time window. Moreover, a shift from an initial first moment of no preference towards either *know* or *think* to a consistent preference towards the agent that thinks was observed for the second time window. We confirmed this result in the third time window, which represents the final moment in the trial when the child was actively prompted to solve the task. In this time window, proportional looking time to the target versus distractor was significantly different from chance, showing a consistent preference towards the distractor element, i.e. the location of the object associated with the lowest degree of certainty. Our exploratory analysis of the time course of target recognition confirmed this finding, showing an increase in fixations towards the agent that thinks, reflected in the negative linear time term of the GLMM.

Taken all together, our results suggest that children as early as 27-months of age exhibit an implicit discrimination between the mental states verbs *know* and *think*. This is in agreement with research supporting that toddlers already show initial understanding of others' knowledge and ignorance (Dunham et al., 2000; Liszkowski et al., 2008; Scott & Baillargeon, 2017, for a review).

The strikingly unexpected finding in our investigation is the directionality of this implicit discrimination: children mostly fixated the agent expressing the lowest degree of certainty, independently on whether the speaking agent in that test phase was the one that knows or thinks. This behaviour led us to consider two alternative explanations: a novelty versus a familiarity effect for the verb *think*.

Firstly, this preference for the thinking agent could reflect a "novelty effect", in which children are confronted with the recurrence of a semantic item they have little experience with (the verb *think*) and are actively engaged trial after trial in learning as much as possible what this new item entails. This is in line with evidence from language acquisition research, which showed correlations between novelty effects and vocabulary proficiency in children (Kormi-Nouri et al., 2005; Marino & Gervain, 2019). In eye-tracking studies, novelty effects have also been widely discussed for their impact on language acquisition (see for a methodological review Blom & Unsworth, 2010; Marno et al., 2016), above all in the context of habituation paradigms, where they are attributed to repetitive exposure time (Houston-Price & Nakai, 2004). In our task, an eventual novelty effect cannot be the simple product of exposure, since children were exposed to the same counterbalanced amount of pseudo-randomized statements from both conditions *know* and *think*. Thus, such novelty effect must reflect children's interest in exploring the yet unknown semantic implications of the newly discovered verb to *think*.

A second, yet contrary hypothesis is that children at this age are not yet acquainted with the mental states interpretation of the verb *know*. So they tend to focus only on what is understandable for their current knowledge of the world during the task. This "familiarity effect" for the thinking agent finds great support in language acquisition literature. While 3-year-olds are not yet able to discriminate *think* and *know* when overtly asked (Montgomery, 1992), one year later children already show prominent production of spontaneous *think* constructions in a sentence recall task in English (Kidd et al., 2010). Furthermore, recent studies showed that parents' use of mental states such as *think* and *understand* predict 2.5-year-old children's performance on anticipatory looking measurements of false belief tasks (see for a meta-analysis Devine & Hughes, 2018; Roby & Scott, 2018). Overall, this suggests that parents' input could already enrich the implicit semantic lexicon of children (see Harris et al., 2005, for the impact of maternal conversation in mental states understanding) to such extent that children exhibit an implicit preference for *think* already at 2.5 years of age. This would be in agreement with the fact that first occurrences in speech for the verb *think* start just a few months later, after their third birthday (Diessel, 2004).

To test both our hypotheses, we correlated the results from the Cognition category of the MSLQ (Mental States Language Questionnaire, adapted to German speakers, see Kristen et al., 2012 for a longitudinal analysis), administered to the parents of the tested children, with the preferential looking score in our eye-tracking task. The correlation between children that already showed active production for the mental verb *know* and that were also more interested in the location associated with the agent that thinks supports the "novelty effect" explanation: the more familiar children are with *know*, the more they focused their attention to the new verb *think*.

Notably, previous research on English speakers showed that children at 27 months of age interpret *know* in their active lexicon rather in the sense of *see* or *recall* than in the epistemic meaning of the verb (Booth et al., 1997). Assuming a similar acquisition pattern can be generalized to German speakers, this would suggest that, as children at 27 months of age are starting to detect and distinguish different degrees of certainty, they are not yet familiar with the usage of the mental states verb *think*, while they already have an initial idea of the scope of the verb *know*. Their preference for the agent in the first and second test phases and the box in the third phase associated with the low-certainty verb led us to speculate that during our eye-tracking task children actively engaged with this new element, learning about the semantic scope of this epistemic state.

Yet we must consider the following limitations when interpreting the findings of this study. Behaviourally, keeping in mind that production does not necessarily involve comprehension, the MSLQ questionnaire revealed to be too broad to investigate in detail the spectrum of possible interpretations that mental states verbs like *know* and *think* can have when used as such a young age. It was therefore impossible to say with certainty that children that are active producers of *know* based on their parents' report are actually already using an epistemic variant of the verb. Based on findings from Booth et al. (1997), 27-month-olds exhibit primarily low levels of meaning of the verb *know*, associated with recall and recognition, rather than understanding and meta-cognition: hence, we are inclined to think that this could have been the case also in our study. A possible confound in the MSLQ results could have been that many children already used *do not know* at a young age, which is often used to express disagreement more than negation of knowledge (Diessel, 2007).

When it comes to the limits of eye-tracking technology, we have to consider that the task might have been too demanding for our sample. Gaze recording implies long sessions sitting still in front of a screen, where toddlers are required to keep 5 to 10 minutes of sustained attention. Through the introduction of pauses, attention grabbers and subsequent recalibration procedures it was possible to record good quality data, truly reflective of cognitive processing in the child's mind. Because of this fragmented recording style, only a careful selection of trials, where children continuously directed their gaze (recorded via an in-sync video camera) to the screen, could be included in the final analyses. This led to a decrease in the total amount of usable data, which are still not as many as we would like in order to derive strong conclusions from our paradigm. On the bright side, the considerable size of our sample and the introduction of a growth curve model, accounting for individual variation across participants as well as for the saliency bias of the speaking agent on screen, contributed to diminishing the impact of these known eye-tracking challenges of developmental research.

Yet, this study has shown that eye-tracking is a powerful tool to help detecting implicit comprehension of epistemic states in early childhood: infants as early as 2.5 years of age show a striking sensitivity for the semantic distinction between *know* and *think*, with a peculiar looking preference for the verb expressing lower certainty.

Due to the novel nature of our task, future studies should validate this unexpected preference for *think* and focus on detangling possible familiarity versus novelty effects, for example by multiple interval arrangement of trials for both preferences (see Houston-Price & Nakai, 2004) . In case of confirmatory evidence, we suggest investigating the origins of this discrimination. This entails shedding further light onto whether the epistemic understanding of low degree of certainty verbs such as *think* precedes the one of higher degrees of certainty; or whether an initial, full-fledged understanding of *know* is a prerequisite for learning about mental states expressing different shades of knowledge of the world. To reach this goal, methods from linguistic research and developmental psychology should be encouraged, as their synergistic use can provide a broader spectrum of insights in the developing mind.

Chapter 5

Main Discussion

5.1 Main Summary

Throughout the work presented here in three experimental studies, we aimed to bring forward methodological improvements to traditional tasks associated with two core cognitive resources necessary for the development of Theory of Mind: false belief understanding and mental states language comprehension. Classic explicit false belief tasks have been at the basis of the theoretical conceptualization that postulates the emergence of a Theory of Mind in children only after four years of age. A growing number of studies criticizes this assumption, proposing that it is simply due to high task requirements that infants younger than four cannot "shine" and display their actual understanding.

Firstly, in study A described here, we aimed to investigate the robustness of findings from a recent behavioural work by Setoh et al. (2016), claiming that children as early as 33 months can pass explicit false belief tasks when low executive processing demands are taken into account. We achieved this goal by conducting a replication effort on a larger sample size, reproducing the original English material in German language, and comparing participants' performance in the task with the results from a traditional explicit false belief representational change task (Hughes & Ensor, 2007).

In our second study B, the goal was to develop a novel ERP paradigm based on a classic behavioural story format task by Wellman & Bartsch (1988), in order to investigate the neural markers of implicit false belief understanding in a task without requirements to actively infer false belief from the agent. Following the same rationale of the first study, we expected that also a reduced processing demands task would elicit correlates compatible with previous implicit false belief processing EEG literature in an adult sample.

In the third and last study C, our aim was to tackle the implicit understanding for mental states language with an eye-tracking study. We postulated that children at 27 months of age already show sensitivity for mental states comprehension, which can be measured by proportional looking times in a semantic discrimination task based on a behavioural paradigm originally developed by Moore et al. (1989).

In order to achieve these objectives, we combined the expertise from classic approaches of developmental psychology to Theory of Mind with techniques used in neuroscience and psycholinguistics (i.e. EEG and eye-tracking), which have the advantage to investigate spontaneous responses evoked by cognitive processes without requiring elicited answers from the participants. In this chapter, we intend to summarise and discuss the key findings of this research effort to cross the borders of our preconceptions towards a common understanding of Theory of Mind at its emergence.

5.2 Main Findings

We present here a summary of the main findings of the three studies.

The behavioural replication of the finding by Setoh et al. (2016) indicated a striking success: within a larger, representative sample of 58 children, 74% answered correctly the final test question in an explicit false belief task where response generation demands were reduced. In contrast, only 5% out of a sample of 42 infants showed full competence in the representational change false belief task by Hughes & Ensor (2007). It is thus not surprising than no statistically significant correlation was found between the performance in the two explicit false belief tasks.

With a novel ERP experiment, we developed a child-friendly paradigm to elicit false belief understanding without overt response from the participant. At the behavioural level, participants showed high accuracy in answering questions targeting true and false belief trials, as well as control questions.

Statistical analysis of neurophysiological data from the same 30 neurotypical adults identified two markers distinguishing false from true belief reasoning: an early positive wave in the left hemisphere at 250 to 400ms, showing temporal similarities with the LPC, and a late wave at 650 to 800ms with reverse polarity between frontal and central-parietal regions, identified as a LSW.

More controversial are the results coming from our new eye-tracking study, where proportional looking patterns of 146 children highlighted no specific preference between the agent that knew (the target) versus the agent that only thought (the distractor) where a desired object was in the first two time windows under analysis. On the contrary, we found a statistically significant above chance preference for the location associated with the distractor verb (*think*) in the final test window.

This finding was in line with the results from a post hoc exploratory general linear mixed-model analysis of the time course of target recognition, which showed that, when controlling for saliency effects, participants consistently preferred the salient agent that thought versus the salient agent that knew the desired object location, across all three time windows. Lastly, a statistically significant positive correlation was found in

27-month-old children which already showed competence in production of the mental verb *know*, based on a parent-administered language questionnaire, and preferential looking towards the agent holding the lowest degree of certainty (*think*).

5.3 Theoretical and Practical Implications

In this section, we discuss theoretical and practical considerations for our understanding of Theory of Mind before the fourth year of age, following our replication of Setoh's task. Furthermore, the role of executive demands, and specifically inhibitory control, as well as mental states language in the development of false belief reasoning is discussed in the light of the findings from our novel ERP and eye-tracking studies. Finally, a broader view on implications and recommendations for developmental and clinical psychology is presented.

5.3.1 Explicit False Belief Reasoning and Inhibitory Control

The findings from our three empirical studies enrich the body of literature on Theory of Mind understanding before the fourth year of age.

In particular, they find their place in a contentious debate sparked in developmental psychology by the different conclusions reached by various paradigms focussing on the attribution of false belief at a young age (Baillargeon et al., 2016; Carruthers, 2013). In the era of the replication crisis in psychology (which affects also false belief studies, see Kulke & Rakoczy, 2018, for a review), controversial conclusions might be ascribed to inconsistent study designs suffering from confounding factors, which make it impossible to clearly assert the nature of the phenomenon under investigation. Moreover, failure to replicate results from original studies has been mainly attributed to questionable research practices (Simmons et al., 2016) and lack of power (Shrout &

Rodgers, 2018).

Taken this into consideration, the confirmatory results of our replication of Setoh et al. (2016) with a large sample bring support to the robustness of the original findings and their interpretation. We can confidently say that already 33-month-old children achieve successful performance in an explicit false belief task, when processing demands are reduced. The floor effect found in our attempt to replicate the representational change false belief task by Hughes & Ensor (2007) can be explained in a similar fashion. Our study included about a third of the participants of the original task and thus it could be missing proper power to detect a true effect (see Maxwell et al., 2015, for a critique on single-study replication attempts). An alternative explanation is that the two false belief tasks, although presumably measuring the same phenomenon, have different solving complexity requirements. This assumption would not be so astonishing, as the representational change task has been previously compared in difficulty to the standard change of location Sally-Anne paradigm (Gopnik & Astington, 1988).

This observation directly leads us to the core issue of the debate: the alleged complexity of the standard false belief task for children below the age of four. A classic view attributing Theory of Mind only after the fourth year of age supports the hypothesis that more advanced representational resources for false belief understanding are necessary to pass traditional false belief tasks, which require explicit answer from the participant (Carruthers, 2013). This interpretation attributes successful performance in the explicit task to a sudden developmental change in the children's mind (Ruffman & Perner, 2005; Saxe et al., 2004), as an intentional switch of perspectives is considered not to be possible in children younger than four (Perner & Roessler, 2012).

In contrast with this view, we presented here evidence of representational resources necessary for mentalising already in 2- to 3-year-old children. These findings are in line with the claim that it is because of the higher cognitive requirements necessary to pass traditional explicit false belief tasks that children younger than four fail (Baillargeon et al., 2016; Bloom & German, 2000). More specifically, it has been proposed that traditional false belief tasks heavily rely on inhibitory control of the correct answer, which must be withheld in order to express an alternative response consistent with one's own representation of another's false belief (Leslie et al., 2004). Furthermore, as the child is requested to verbally answer a standard question, these tasks could activate response generation processes in the brain that are in competition with the concurrent information processing resources needed for the child to reason about the mental states of other agents, represent their knowledge of the world, and ultimately infer their future behaviour (Setoh et al., 2016). The Setoh's task overcomes these two limitations by introducing practice trials not related to inference of beliefs, which help reducing response generation demands, so not to overwhelm infants' overall cognitive processing system, and in turn allow children to handle the processing demands of false belief representation.

The confirmatory outcome of our study, i.e., that reducing inhibitory control demands allows children even below the age of three to pass an explicit false belief task, is in agreement with the interpretation of other successful studies on infants which employed low processing variants of the traditional Sally-Anne scenario (Bloom & German, 2000; Carpenter et al., 2002). Further evidence in support of this hypothesis comes from spontaneous false belief tasks tackling implicit understanding, where inhibitory control demands are reduced, because no answer has to be verbally elicited by the child to solve the task correctly. For instance, positive findings before the fourth year of age have been recorded with various spontaneous responses, such as violation of expectation (Onishi & Baillargeon, 2005) as well as preferential and anticipatory looking (Scott et al., 2012; Southgate et al., 2007, among others). Other tasks that showed positive performance early on are elicited intervention tasks, where children were able to succeed in retrieving an object for an agent based on their false belief (Buttelmann et al., 2009). Overall, spontaneous implicit as well as explicit tasks with reduced processing demands support the existence of some primordial capacities for Theory of Mind reasoning, available in the infant's mind earlier than previously thought.

A relevant theoretical consideration follows the evidence collected by these studies: rather than a fundamental change in children's false belief representation at the age of four, there might be a continuity from implicit to explicit false belief understanding between infancy and early childhood. This conceptual continuity hypothesis (Baillargeon et al., 2010) suggests the presence of a unique psychological system reasoning about mental states, accessible from infancy onward, which is affected by developmental progression of low and high order cognition (Baillargeon et al., 2016; Leslie et al., 2004). In this view, only when additional cognitive processing resources are available throughout development of key neural regions dedicated to mentalising processes, such as the prefrontal cortex (Moriguchi et al., 2007; Grosse Wiesmann et al., 2017), traditional false belief tasks with higher inhibitory demands become manageable for children. This is also in alignment with the results coming from our EEG study, which hints to the effect of neural ERP components associated with false belief reasoning to be strongly localized in the prefrontal (left and midline) recording sites.

As a practical implication, it follows that, to be able to investigate the emergence of Theory of Mind in the infant brain, false belief tasks tackling essential representational resources for mentalising need to be reduced in complexity to avoid cognitive resources competition that obscures real competences.

5.3.2 Processing Demands in Implicit False Belief Tasks

Our findings on the implicit understanding of false belief in a novel paradigm with reduced processing demands find their place against this background. Two distinct temporal markers emerged during the EEG task, identified by a divergence in neurophysiological amplitude of responses during the target event. The presence of consistent neural correlates supports the task ability to detect and elicit discrimination between true and false belief conditions, which was our main goal while developing a new false belief ERP paradigm. Most interestingly, in agreement with our second hypothesis, the two waveforms showed similarities with previous LPC (Cao et al., 2012; Meinhardt et al., 2011; Zhang et al., 2009) and LWS found in false belief literature (Geangu et al., 2012; Meinhardt et al., 2011; Sabbagh & Taylor, 2000). This outcome suggests that passive listening to the elicited belief of the agent is sufficient in neurotypical adults to instantiate the same inferential reasoning process present in classic false belief tasks with higher processing requirements.

This proposition is encouraging in two ways. On one side, the fact that a reduced processing demands task is able to elicit consistent markers, with resemblance to components previously found in literature, give us confidence in the validity and suitability of the novel paradigm to investigate effectively implicit false belief understanding.

On the other side, the temporal specificity of these neural markers is in line with previous research that suggests that early LPC and late LSW represent the neurophysiological manifestation of two different neural processing events necessary for false belief reasoning (Meinhardt et al., 2011). Following this hypothesis, the broad LPC waveform might be evoked by generalized mismatch (similar to a P3a component, see Polich, 2007) between the subject own knowledge of reality and the incoming false belief information (Zhang et al., 2009). This initial moment is followed by the LSW, as specific neural correlate of the decoupling process distinguishing mental states from reality, necessary for false belief reasoning to solve the task at hand (Liu et al., 2009b). Additionally, using source analysis, this and a previous ERP study found LSW activity to be localized in the prefrontal cortex in both adults and children able to pass a false belief task (Liu et al., 2004). This finding is in agreement with the hypothesis that maturation of the prefrontal cortex is necessary for the development and adult-like employment of Theory of Mind capacities (Grosse Wiesmann et al., 2017). It has also been previously proposed that LSW reflects difficulty of operations of verbal or conceptual working memory from different domains of information (Lang et al., 1987; Mecklinger & Pfeifer, 1996), hence, LSW in belief reasoning tasks could reflect the social-cognitive inferential processing in working memory necessary to solve the mentalising problem at hand (Liu et al., 2009b).

Such complexity in attributing a unique interpretation to neural markers of implicit false belief tasks is comprehensible in light of the multiple, competing cognitive abilities which are involved in the process of Theory of Mind understanding: from working memory to executive functions and language abilities (Astington & Jenkins, 1999). It is not surprising that, with so many resources needed simultaneously, the cognitive processing system of young children results in being overloaded, as inhibitory control of infants is still immature (Diamond, 2013).

While it is impossible to have complete control on the impact of concurrent cognitive resources while administering and performing an implicit false belief task, our ERP paradigm shows high potential to be effectively employed for investigating the temporal specificity of false belief understanding in the infant population. This is particularly due to its short story format, lighter on working memory requirements, and the lack of explicit false belief inferential requests, two aspects that lead to overall reduced processing demands for solving the false belief task.

Taken together, this investigation points to the fact that performance in explicit as well as implicit false belief tasks is complex to assess, as it can depend on other cognitive resources, such as processing speed, working memory, inhibitory control, and language abilities (Carlson & Moses, 2001; Chasiotis et al., 2006; Milligan et al., 2007).

As brilliantly stated by Setoh et al. (2016), false belief understanding is an essential component, but it is only one of the requirements for the success at false belief paradigms, which depends on the full range of processing demands of the task.

A potential implication is that concurrent tasks with reduced cognitive demands should be employed to tackle and understand the interplay between different processing resources involved in mentalising. For instance, success in traditional false belief tasks correlates with performance in conflict inhibitory control tasks (Carlson & Moses, 2001; Chasiotis et al. 2006). This entails focussing our research efforts not only on spontaneous false belief tasks, but also on assessing executive functions, e.g. by assessing recognition memory and processing speed in toddlers (see Rose et al., 2002; Rose et al., 2004, for an eye-tracking battery developed for this purpose).

5.3.3 The Interplay of Language and Theory of Mind

Beside executive resources, enduring controversy still surrounds the role of language in the development of a full-fledged Theory of Mind. Studies suggest that language abilities develops in parallel and can be fostered by social cognition and mentalising (Harris et al., 2005), while others assume language to be a necessary precursor to the development of false belief understanding (Astington & Jenkins, 1999; Ruffman et al., 2002). In order to provide effectively an empirical answer to this theoretical question, it seems first necessary to implement language comprehension tasks that do not rely on eliciting verbal responses from toddlers and infants, as they could be not representative of actual competence, due to their limited processing capacities discussed above.

Albeit with some reservations (see section 5.4 on Limitations and Future Research), our newly developed eye-tracking paradigm for investigating early sensitivity to semantic difference of epistemic verbs revealed to be suitable for young children. From a technical point of view, less than 10 percent of the infants involved in the task dropped out due to task-related issues and a noticeable number of valid fixations was recorded up to the last test phase, suggesting that infants were engaged long enough to keep sustained attention throughout the whole sequence. Thus, the task does not seem to be too heavy in concurrent processing load for 2.5-year-old children. A special mention regards the saliency effect of the agents on screen (irrespectively of their degree of knowledge), which affects inhibitory resources and should be addressed in further studies.

Hence, once controlled for the saliency effect of the moving agents, the striking finding of this study was a steady preference of 27-month-old children for the agent holding, and the location associated to, the lowest degree of certainty. On one side, such remarkable difference in proportional looking times can be taken as indication that children below 3 years of age already have an early sensitivity for discriminating mental states language. This is in line with previous research showing that infants already have initial comprehension of others' knowledge and ignorance (Liszkowski et al., 2008; Scott & Baillargeon, 2017, for a review). On the other side, the directionality of such sensitivity raises questions in terms of its interpretability. Based on the correlation analyses of proportional looking times to the distractor (*think*) with epistemic verbs production (*know*) from a mental states language questionnaire administered to parents, we found support for a novelty effect explanation, which has been previously shown to be related to the level of vocabulary proficiency in children (Kormi-Nouri et al., 2005; Marino & Gervain, 2019).

Following this novelty effect interpretation, children who already show productive competence of the epistemic verb *know* engage actively with, and exhibit a strong preference for, the distractor verb *think*, giving us a sneak peek in the unravelling of the implicit processes of language acquisition. This finding is in line with previous linguistic research claiming that 27-month-olds have already a basic understanding of knowing, although they interpret this verb mainly in the semantic meaning of recognition or, when preceded by negation, disagreement, rather than in the meaning of understanding and knowledge (Booth et al., 1997). It also supports the view that at 2.5 years of age children have not yet grasped the epistemic nuances of the conceptual verb to *think*, but are intrigued by learning its semantic applications. This novelty effect is also in agreement with research on child-directed speech, which found that most exposure to mental states vocabulary, and particularly to *think*, does not involve contextual situations where an agent has a false belief (Diessel & Tomasello, 2001; Hacquard & Lidz, 2019).

Although our mental states language task alone cannot provide an answer to the question whether language abilities are a precursor of Theory of Mind or not, we can make some speculations. Our task revealed to be appropriate in detecting early sensitivity already in 2.5-year-old children. If mental states language were indeed essential for mentalising, we would expect to find a positive correlation between above chance preferential discrimination of mental states verbs at 2.5 years of age and successful performance in an explicit false belief task with reduced inhibitory demands, such as the Setoh's task, at 3 years of age. This correlation would be particularly strong for children that are active users of the epistemic verb to *know* already at a young age.

We would also expect young children exhibiting higher performance in executive functions tasks, specifically in inhibitory control and working memory, to evoke similar neural correlates of implicit false belief understanding as found by our ERP task with reduced processing demands in adults. These children could in fact have an advantage in freeing cognitive processing resources to dedicate to false belief reasoning even before the fourth year of age. Such findings would not only be in line with research supporting language and executive functions as necessary for false belief reasoning (German & Leslie, 2000; Milligan et al., 2007), but would also provide support for the conceptual continuity hypothesis (Baillargeon et al., 2010), in a framework which sees the emergence of Theory of Mind as the results of the interplay of many concurrent cognitive resources.

5.3.4 Developmental and Clinical Implications

All in all, we see a need for developmental psychology to focus on increasing replications, with larger sample sizes and targeted age groups, of core studies whose findings are at the basis of current theories in social cognition. Particular attention should be dedicated to develop tasks suitable for toddlers and infants with reduced extraneous cognitive load in order not to underestimate performance due too high executive function demands. Longitudinal studies should be preferred, in order to detect and define developmental patterns from infancy to childhood via comparison of elicited behavioural and spontaneous responses to false belief and language comprehension tasks. In turn, findings from these carefully designed studies would lead to refine the theoretical basis of the onset of Theory of Mind and its relation to concurrently developing cognitive resources.

Such research effort can have striking societal implications for early uncovering of neuropathologies and psychopathologies, particularly if specific representational resources which are needed for a full-fledged development of Theory of Mind are impaired. The detection of concomitant signs pointing to below average or inadequate performance in executive functions, mental states language, and false belief reasoning, could be of use to support diagnostic tools, as well as early intervention for mental disorders, which starts in early childhood and affect the sphere of social cognition.

In this regard, autism spectrum disorder is surely the most studied neurodevelopmental condition affected by a deficit in the ability to infer mental states of others (Brüne & Brüne-Cohrs, 2006, for a review) as well as in executive function, two aspects that have been debated to be causally related or independent from each other (Baron-Cohen & Swettenham, 1997; Perner & Lang, 1999). Thus, tasks testing Theory of

Mind abilities have the potential to become an aid to early diagnosis, when controlled for the role of executive processing demands (Charman, 2000).

Observed deficits in Theory of Mind task performance have been shown, when matched for general cognition, also in frontotemporal dementia, but not in Alzheimer's disease (Bora et al., 2015 for a meta-analysis). Interestingly, in a previous study comparing the two populations, the Alzheimer's group failed only one false belief task, which placed heavy demands on working memory (Gregory et al., 2002).

A tight connection between Theory of Mind and executive functions clearly emerged also for brain-damaged patients tested with a non-linguistic belief reasoning task with reduced executive demands. Patients showed overall Theory of Mind deficits, but prefrontal cortex lesions were associated to belief reasoning errors associate with executive function problems, while patients with temporo-parietal junction lesions made representational errors, in supports of a spatial specificity of Theory of Mind neural processes (Apperly et al., 2004).

Another neuropsychiatric group affected by inability to appropriately infer beliefs and intentions of others is schizophrenic patients. A combination of behavioural and neuroimaging studies suggested that, compared to typically developed controls, these patients exhibit also prefrontal cortical reductions associated with concomitant poor performance in Theory of Mind (Hirao et al., 2008). It is important to notice that all these studies looking for links between neuropsychological impairments and Theory of Mind used different batteries, also leading to contrasting results or lack thereof (Abdel-Hamid et al., 2009).

Interestingly, using a video-based Movie for the Assessment of Social Cognition, a significantly correlation between hypomanic episodes and under-mentalising abilities was found in bipolar patients (Montag et al., 2010). And by using the same test, Sjølie et al. (2020) found an association between Theory of Mind and intelligence, speed of processing, verbal or visual memory, and non verbal working memory in the schizophrenic population.

This brief overview of clinical investigations undoubtedly shows that there is much still unknown about the implications of Theory of Mind impairments on psychopathologies during adulthood. Due to the lack of standardization in the tests used across studies, it is difficult to associate different disorders with a deficit in a specific underlying mentalising ability. Furthermore, we cannot stress enough the necessity to include test batteries to control for executive functions and language skills when performing Theory of Mind tasks, as much as with children as with neurodivergent populations, which can suffer from independent impairments to executive control or linguistic abilities.

A causal link between developmental and clinical research is not unexpected. As Theory of Mind development is initially fostered by parent-child interactions and reflections on child's mental states by the parent, personality disorders associated with attachment disturbances have also been recently evaluated against Theory of Mind deficits. For instance, adolescents with high level of borderline pathology showed attachment-based reflective disturbances strongly related to Theory of Mind impairments (Vanwoerden et al., 2019). As a caveat, we need to be careful when discussing the relation between ability to attribute mental states and social impairments, as there is not yet any verified casual relation between the two. For example, Asperger individuals have been found to show cognitive Theory of Mind abilities (based on the classic false belief task by Baron-Cohen et al., 1985) but, despite these subjects' knowledge about others' mental states, they were unable to apply this knowledge effectively in social settings (Bowler et al., 1992).

This finding implies that theoretical competences are not sufficient alone to solve social deficits, and more all-around approaches that include an integration of mentalising tasks with non-social cognition (e.g. cognitive remediation therapy, used for autism spectrum disorder interventions, see Fernandez-Sotos et al., 2019) should be preferred. Targeted Theory of Mind training could support psychotherapeutically young populations at risk of personality or psychopathological disorders. Two examples in this direction combined cognitive remediation and Theory of Mind trainings, showing maintained social well-being and functional improvements in schizophrenic participants after one year (Thibaudeau et al., 2017) and more recently even after a 3 years period (Bechi et al., 2020).

Ultimately, it is our aspiration that once the relation between the emergence of Theory of Mind, executive functions, and language abilities is clarified, the scientific community will be able to focus on developing target combined programs to support even young children with social cognition deficits in developing a functional Theory of Mind network and live at best in society.

5.4 Limitations and Future Research

As a project focussing on methodological improvement and development of new paradigms suitable for developmental research, the three empirical studies presented here are not exempt from limitations.

In the replication of Setoh et al. (2016), we could not rule out language specific effects that could have facilitated German-speaking toddlers in solving the false belief scenario, compared to the original English speaking sample. Although possible this is highly unlikely, as previous studies on early Theory of Mind performance did not find cross-cultural effects on false belief reasoning (Chasiotis et al., 2006). In addition, the inability to replicate the findings from Hughes and Ensor (2007) made it impossible to assess the validity of the representational change task in measuring false belief understanding. As previously mentioned, this failure could be explained by the reduced sample size of our sample compared to the original study, but it does not provide an explanation for the lack of any statistical correlation between tasks, since based on our sample size our study would have had enough power to detect an effect between the two tasks.

Multiple labs replications and ultimately meta-analysis are highly advised to overcome experimental setup differences and small samples when performing replication studies (Maxwell et al., 2015). Thus, we strongly support other international labs to implement their own version of the Setoh et al.'s and Hughes and Ensor's paradigms to validate these behavioural findings and remove some of the scepticism around explicit false belief evidence collected before the fourth year of age. Due to its quick setup and execution, low inhibitory demands, and robust performance, we strongly encourage the Setoh et al.'s paradigm to be included in the future as a standard explicit false belief task for Theory of Mind batteries.

Some of the traditional pitfalls of neurophysiological studies have already been presented in the Main Introduction (Chapter 1). In the new EEG study described here, although the overall accuracy was very high, we did not accomplish to reach ceiling performance in the control questions and false versus true belief behavioural response of our participants. Clearer instructions avoiding psychological biases as well as more breaks between the EEG sessions could help to sort out this fallacy.

When looking into ERP analyses, the 150ms time latency of the LSW detected in our study tends to be shorter than what found in literature for late evoked components. A switch to opposite polarities at different axial electrodes in the grid analysis (Liu et al, 2009b) made a homogeneous pull of the data impossible (as it would have overall net neutralizing effects on the differential areas of the ERP waveforms). On the other side, the switch of polarities could represent a dipole generator located between the frontal and centro-parietal axis and, in such case, it would have required a change of referencing system to be avoided. Despite that, we tend to find the dipole possibility unlikely, as the peak activity at frontal and parietal sites did not exhibit identical latencies.

In this regard, current best EEG analysis practices in cognitive neuroscience suggest to avoid electrode sites selection for quantifying components based on previous positive findings from literature (Luck & Gaspelin, 2017). In addition, visual inspection of grand-averaged data to select time windows should be deterred as this procedure inflates the possibility of finding spurious effects. As the lateralization effect found for LPC could potentially have this origin, bootstrapping analyses procedures (see Maris & Oostenveld, 2007; Pernet et al., 2015) should be preferred in the future to identify relevant electrode locations (spatial clusters) and time windows (temporal cluster) at which conditions differ to validate this finding. It is worth noticing that only very recently these advancements in EEG-methodology have reached the developmental field, with the scope to help maximising reproducibility via cluster-based permutations and report of effect sizes (Meyer et al., 2021).

Furthermore, the employed bins analysis procedure (Gomarus, 2010; Müller et al., 2016) should be avoided for time dependent data, as the temporal evolution of ERP signals makes it unlikely for consecutive segments to be uncorrelated. Instead, for future studies implemented with the intent to replicate this paradigm, we suggest to focus on frequency-based analysis, which has been shown to have the advantage of dissociating hard-to-distinguish language related ERP components (Roehm et al., 2004). Finally, a powerful tool for investigating the spatial resolution of time-locked components is to apply source localization algorithms based on high-density EEG systems, which would allow to reconstruct the localization of effects to understand the underling neural structures involved in false belief processing (Michel & Brunet, 2019).

When it comes to our novel eye-tracking paradigm on epistemic language, we particularly focused on having a representative sample size to increase the power to detect spontaneous correlates of language sensitivity in young children. Instead, the main limitation of our mental states language comprehension task regards its story format, in which the two agents alternate their knowledge statements about the possible location of a hidden object. In order to make the necessary trial repetition procedure interesting, we added a jiggling sound followed by a brief movement of the agent towards the chosen location to the videos. Such element ensured sustained attention in our 27-month-old sample, but also added a bias towards the moving element on screen, as reorienting of attention was too difficult to inhibit for young children. Controlling a posteriori for saliency required advanced statistical data analyses at the expense of an intuitive interpretability of results.

Future studies that rely on such paradigm, which showed high potential to detect discriminatory sensitivity for lexical elements in young children, should focus on lowering saliency of the involved agents across the storyline, and shortening trials length. Any improvement to the task procedure must be considered mindfully, as the major challenge of keeping the participant engaged long enough to collect sufficient high quality data in spontaneous response tasks with infants holds true here. Once such technicalities have been addresses, the task format would be beneficial to other research questions relevant to language comprehension in early childhood.

Broadly speaking, the scope of our three studies was to shed light onto processing capacities for false belief understanding and mental states language comprehension before the fourth year of age. What escapes the limits of our paradigms is whether it is indeed possible to pinpoint an exact age at which children first have the necessary representational resources, supported by concomitant executive functions, to put themselves in someone else perspective and infer their actions based on understanding of others' mental states.

Furthermore, it is yet unknown whether implicit false belief understanding is an independent cognitive capacity, which develops in parallel to linguistic competences and executive functions, or whether, without these other essential cognitive abilities, no full-fledged Theory of Mind reasoning is possible altogether.

It remains unclear what is the exact role of specific executive functions in limiting performance of young children: in this dissertation we have seen examples of inhibitory control reduction by response inhibition (see replication of Setoh et al., 2016), by removal of inferential processing (in our neural correlates of false belief task), and by working memory demands (particularly relevant in our eye-tracking task). Not only separating individual effects of concurrent processing resources necessary to solve the task remains challenging, but results from Theory of Mind research show inconsistent findings in this regard. Strikingly, populations that exhibit enhanced inhibitory control, such as native bilingual children, showed no benefit in traditional false belief tasks (Kovács, 2009).

Eventual cross-cultural differences in developmental patterns of Theory of Mind have also to be taken into account in future research. For instance, while controlling for language understanding abilities, German, Costa Rican, and Cameroonian preschoolers exhibited a culture-independent relation between inhibitory control and false belief understanding (Chasiotis et al., 2006). On the contrary, a following massive study on Chinese preschoolers showed reversal in performance on two traditional high inhibition false belief tasks by Wellman & Liu (2004) compared to the original sample, and successful outcomes could be predicted by individual differences in working memory, but not conflict inhibition (Duh et al., 2016).

Finally, in order to get more insights on the potential role of language as precursor of Theory of Mind abilities, it would be meaningful to investigate whether performance deficits in non-linguistic false belief tasks are visible in children showing semantic language impairments, as well as adults affected by neurolinguistics brain damages as, e.g., Wernicke's aphasia.

To reach these ambitious goals international longitudinal studies with batteries for implicit and explicit false belief understanding with reduced processing demands, testing in parallel for the development of executive functions and mental states language comprehension, should be favoured. When possible, new technologies that can give us insights without explicit answers from young participants, such as eye-tracking, fNIRS, and EEG, should be included to discriminate spontaneous correlates of mentalising abilities.

Ultimately, an exciting and ambitious challenge will consist in monitoring the brain development of key areas, such as the medial prefrontal cortex and the right temporo-parietal junction, by neuroimaging techniques. This will allow to longitudinally track the development of neural patterns of Theory of Mind processing in neurotypical and neurodivergent populations, as well as to investigate possible long lasting effects of combined Theory of Mind and cognitive training on cortical neuroplasticity.

5.5 Conclusion

In this dissertation, we presented methodological improvements in three paradigms, which have the potential to bring forward our understanding of Theory of Mind at its emergence.

Thanks to the replication of Setoh et al. (2016), we brought evidence towards a conceptual continuity view of Theory of Mind development, claiming that small infants have a primordial understanding of others' people mental states, which develops continuously across childhood. As this explicit false belief task relies on lowering processing demands to show successful performance, here we support the hypothesis that it is a lack of inhibitory control and an overload in cognitive resources required for eliciting correct response generation that hinder children younger than four in classic false belief tasks. Hence, it comes without saying that representational resources have to be available to the child mentalising system to be able to appreciate others' epistemic differences.

In this regard, the findings of the novel eye-tracking task, developed to investigate early comprehension of mental states verbs, presented evidence of early sensitivity to *know* versus *think* discrimination already in 2.5-year-old infants. Our outcomes need to be corroborated by methodological improvements in reducing stimuli saliency, but they pave the way to further attempts to characterize children comprehension via their spontaneous responses, instead of relying to parents' report or contextual linguistic production. Such early mental states understanding is in our view necessary along the path to a full-fledged Theory of Mind: a fascinating hypothesis that could be investigated by examining within the same study design the relationship between semantic language acquisition and the development of false belief understanding.

Ultimately, the novel implicit false belief ERP task contributes to this goal, providing a tool that, in virtue of the reduced processing demands and without necessity of inferential reasoning of others' belief, can carry valid evidence of neural markers elicited by false belief understanding, potentially even in young children. It will certainly be an exciting task to apply this paradigm in longitudinal studies from infancy into adulthood exploring the implicit and explicit correlates of Theory of Mind with other core cognitive abilities that could be precursors for its development, such as executive functions and mental states language.

In conclusion, this dissertation brings forward the field in two ways: firstly, it contributes to enforce a sense of robustness in the long-standing effort of developmental researchers committed to the difficult task of unravelling mentalising processes, by confirming previous controversial but extraordinary findings on explicit false belief in early childhood. Secondly, it presents the undoubted benefits of combining traditional behavioural research paradigms, which are at the basis of psychological theories on child development, with methodologies from neuroscience and psycholinguistics, in order to shed light onto the nature of early Theory of Mind capacities.

This inspiring goal presents unquestionable challenges but also an exciting opportunity: to advance a uniform vision on our understanding of core social cognitive mechanisms developing in that marvellous black box which is the human mind.

Bibliography

Abdel-Hamid, M., Lehmkämper, C., Sonntag, C., Juckel, G., Daum, I., & Brüne, M. (2009). Theory of mind in schizophrenia: the role of clinical symptomatology and neurocognition in understanding other people's thoughts and intentions. *Psychiatry Research, 165*(1-2), 19-26.

Aichhorn, M., Perner, J., Kronbichler, M., Staffen, W., & Ladurner, G. (2006). Do visual perspective tasks need theory of mind?. *Neuroimage, 30*(3), 1059-1068.

Ali, F., & Chamorro-Premuzic, T. (2010). Investigating theory of mind deficits in nonclinical psychopathy and Machiavellianism. *Personality and Individual Differences, 49*(3), 169-174.

Alighieri, D. (2009). *The Divine Comedy* (H. F. Cary, Trans.). Wordsworth Editions.

Antonietti, A., Liverta-Sempio, O., Marchetti, A., & Astington, J. W. (2006). Mental language and understanding of epistemic and emotional mental states. In A. Antonietti, O. Liverta-Sempio, & A. Marchetti (Eds.), *Theory of mind and language in developmental contexts.* (pp. 1-30). Springer Science & Business Media.

Apperly, I. A., Samson, D., Chiavarino, C., & Humphreys, G. W. (2004). Frontal and temporo-parietal lobe contributions to theory of mind: neuropsychological evidence from a false-belief task with reduced language and executive demands. *Journal of Cognitive Neuroscience, 16*(10), 1773-1784.

Arguel, A., & Jamet, E. (2009). Using video and static pictures to improve learning of procedural contents.*Computers in Human Behavior, 25*(2), 354-359.

Astington, J. W., & Jenkins, J. M. (1999). A longitudinal study of the relation between language and theory-of-mind development. *Developmental Psychology, 35*(5), 1311.

Azarian, R. (2010). Social ties: Elements of a substantive conceptualization. *Acta Sociologica, 53*(4), 323-338.

Baillargeon, R., Buttelmann, D., & Southgate, V. (2018). Invited commentary: Interpreting failed replications of early false-belief findings: Methodological and theoretical considerations. *Cognitive Development, 46*, 112-124.

Baillargeon, R., Scott, R. M., & Bian, L. (2016). Psychological reasoning in infancy. *Annual Review of Psychology, 67*(1), 159-186.

Baillargeon, R., Scott, R. M., & He, Z. (2010). False-belief understanding in infants. *Trends in Cognitive Sciences, 14*(3), 110–118. https://doi.org/10.1016/j.tics.2009.12.006

Baron-Cohen, S. (2000). Theory of mind and autism: A review. *International Review of Research in Mental Retardation, 23*, 169-184.

Baron-Cohen, S., & Swettenham, J. (1997). Theory of mind in autism: Its relationship to executive function and central coherence. *Handbook of Autism and Pervasive Developmental Disorders, 2*.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"?. *Cognition, 21*(1), 37-46.

Bartsch K., Wellman H. M. (1995). *Children talk about the mind*. Oxford University Press.

Bassano, D. (1996). Functional and formal constraints on the emergence of epistemic modality: A longitudinal study on French. *First Language, 16*(46), 077–113.

https://doi.org/10.1177/014272379601604605

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1-48. https://doi.org/10.18637/jss.v067.i01

Bechi, M., Bosia, M., Buonocore, M., Agostoni, G., Bosinelli, F., Silvestri, M. P., ... & Cavallaro, R. (2020). Stability and generalization of combined theory of mind and cognitive remediation interventions in schizophrenia: Follow-up results. *Psychiatric Rehabilitation Journal, 43*(2), 140.

Blair, J., Sellars, C., Strickland, I., Clark, F., Williams, A., Smith, M., & Jones, L. (1996). Theory of mind in the psychopath. *Journal of Forensic Psychiatry, 7*(1), 15-25.

Blom, E., & Unsworth, S. (2010). *Experimental methods in language acquisition research.* John Benjamins Publishing.

Bloom, P., & German, T. P. (2000). Two reasons to abandon the false belief task as a test of theory of mind. *Cognition, 77*(1), B25-B31.

Booth, J. R., Hall, W. S., Robison, G. C., & Kim, S. Y. (1997). Acquisition of the mental state verb know by 2-to 5-year-old children. *Journal of Psycholinguistic Research, 26*(6), 581-603.

Bora, E., Walterfang, M., & Velakoulis, D. (2015). Theory of mind in behavioural-variant frontotemporal dementia and Alzheimer's disease: a meta-analysis. *Journal of Neurology, Neurosurgery & Psychiatry, 86*(7), 714-719.

Bowler, D. M. (1992). "Theory of Mind" in Asperger's Syndrome Dermot M. Bowler. *Journal of Child Psychology and Psychiatry, 33*(5), 877-893.

Bowman, L. C., Liu, D., Meltzoff, A. N., & Wellman, H. M. (2012). Neural correlates of belief-and desire-reasoning in 7-and 8-year-old children: an event-related potential study. *Developmental Science, 15*(5), 618-632.

Bretherton, I., & Beeghly, M. (1982). Talking about internal states: The acquisition of an explicit theory of mind. *Developmental Psychology, 18*(6), 906.

Brooks, R., & Meltzoff, A. N. (2015). Connecting the dots from infancy to childhood: A longitudinal study connecting gaze following, language, and explicit theory of mind. *Journal of Experimental Child Psychology, 130*, 67-78.

Brüne, M. (2005). "Theory of mind" in schizophrenia: a review of the literature. *Schizophrenia Bulletin, 31*(1), 21-42.

Brüne, M., & Brüne-Cohrs, U. (2006). Theory of mind—evolution, ontogeny, brain mechanisms and psychopathology. *Neuroscience & Biobehavioral Reviews, 30*(4), 437-455.

Buttelmann, D., Carpenter, M., & Tomasello, M. (2009). Eighteen-month-old infants show false belief understanding in an active helping paradigm. *Cognition, 112*(2), 337-342.

Cao, B., Li, Y., Li, F., & Li, H. (2012). Electrophysiological difference between mental state decoding and mental state reasoning. *Brain Research, 1464*, 53-60.

Carlson, S. M., & Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development, 72*(4), 1032-1053.

Carlson, S. M., Mandell, D. J., & Williams, L. (2004). Executive function and theory of mind: stability and prediction from ages 2 to 3. *Developmental Psychology, 40*(6), 1105.

Carlson, S. M., Moses, L. J., & Breton, C. (2002). How specific is the relation between executive function and theory of mind? Contributions of inhibitory control and working memory. *Infant and Child Development: An International Journal of Research and Practice, 11*(2), 73-92.

Carpenter, M., Akhtar, N., & Tomasello, M. (1998). Fourteen-through 18-month-old infants differentially imitate intentional and accidental actions. *Infant Behavior and Development, 21*(2), 315-330.

Carpenter, M., Call, J., & Tomasello, M. (2002). A new false belief test for 36-month-olds. *British Journal of Developmental Psychology, 20*(3), 393-420.

Carpenter, M., Nagell, K., & Tomasello, M. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development, 63*(4, Serial No. 255)

Conklin, K., & Pellicer-Sánchez, A. (2016). Using eye-tracking in applied linguistics and second language research. *Second Language Research, 32*(3), 453–467. https://doi.org/10.1177/0267658316637401

Carruthers, P. (2013). Mindreading in infancy. *Mind & Language, 28*(2), 141-172.

Charman, T. (2000). Theory of mind and the early diagnosis of autism. In S. Baron-Cohen, H. Tager-Flusberg, & D. J. Cohen (Eds.), *Understanding other minds: Perspectives from developmental cognitive neuroscience* (pp. 422–441). Oxford University Press.

Chasiotis, A., Kiessling, F., Hofer, J., & Campos, D. (2006). Theory of mind and inhibitory control in three cultures: Conflict inhibition predicts false belief understanding in Germany, Costa Rica and Cameroon. *International Journal of Behavioral Development, 30*(3), 249-260.

Clements, W. A., & Perner, J. (1994). Implicit understanding of belief. *Cognitive Development, 9*(4), 377-395.

Copeland, L., & Gedeon, T. (2013, December 2-5). *Measuring reading comprehension using eye movements* [Paper presentation]. 2013 IEEE 4th International Conference on Cognitive Infocommunications (CogInfoCom), Budapest, Hungary. https://doi.org/10.1109/CogInfoCom.2013.6719207

De Villiers, J. G., & Pyers, J. E. (2002). Complements to cognition: A longitudinal study of the relationship between complex syntax and false-belief-understanding. *Cognitive Development, 17*(1), 1037-1060.

De Villiers, J. G., & De Villiers, P. A. (2000). Linguistic determinism and the understanding of false. *Children's Reasoning and the Mind, 191.*

Devine, R. T., & Hughes, C. (2018). Family correlates of false belief understanding in early childhood: A meta-analysis. *Child Development, 89*(3), 971-987.

Diamond, A. (2013). Executive functions. *Annual Review of Psychology, 64*, 135.

Diessel, H. (2004). *The acquisition of complex sentences* (Vol. 105). Cambridge University Press.

Diessel, H. (2007). Frequency effects in language acquisition, language use, and diachronic change. *New Ideas in Psychology, 25*(2), 108-127.

Diessel, H. & Tomasello, M. (2001). The acquisition of finite complement clauses in English: A corpus-based analysis. *Cognitive Linguistics, 12*(2), 97-142. https://doi.org/10.1515/cogl.12.2.97

Dolan, M., & Fullam, R. (2004). Theory of mind and mentalizing ability in antisocial personality disorders with and without psychopathy. *Psychological Medicine, 34*(6), 1093-1102.

Duh, S., Paik, J. H., Miller, P. H., Gluck, S. C., Li, H., & Himelfarb, I. (2016). Theory of mind and executive function in Chinese preschool children. *Developmental Psychology, 52*(4), 582.

Dunham, P., Dunham, F., & O'Keefe, C. (2000). Two-year-olds' sensitivity to a parent's knowledge state: Mind reading or contextual cues? *British Journal of Developmental Psychology, 18*(4), 519-532.

Fernandez-Sotos, P., Torio, I., Fernandez-Caballero, A., Navarro, E., Gonzalez, P., Dompablo, M., & Rodriguez-Jimenez, R. (2019). Social cognition remediation interventions: A systematic mapping review. *PloS One, 14*(6), e0218720.

Flavell, J. H., Green, F. L. & Flavell, E. R. Developmental changes in young children's knowledge about the mind. *Cognitive Development, 5*, 1–27 (1990).

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences, 6*(2), 78-84.

Frith, C. D., & Frith, U. (2007). Social cognition in humans. *Current Biology, 17* (16), 724-732.

Frith, U. (1994). Autism and theory of mind in everyday life. *Social Development, 3*(2), 108-124.

Geangu, E., Gibson, A., Kaduk, K., & Reid, V. M. (2012). The neural correlates of passively viewed sequences of true and false beliefs. *Social Cognitive and Affective Neuroscience, 8*(4), 432-437.

German, T. P., & Leslie, A. M. (2000). Attending to and learning about mental states. *Children's Reasoning and the Mind*, 229-252.

Gomarus, H. K. (2010). *The psychophysiology of selective attention and working memory in children with PPDNOS and/or ADHD.* University of Groningen.

Gómez, J. C. (2009). *Apes, monkeys, children, and the growth of mind.* Harvard University Press.

Gopnik, A., & Astington, J. W. (1988). Children's understanding of representational change and its relation to the understanding of false belief and the appearance-reality distinction. *Child Development*, 26–37.

Gratton, G., Coles, M.G., Donchin, E., 1983. A new method for off-line removal of ocular artifacts. *Electroencephalography and Clinical Neurophysiology, 55*, 468–484.

Gregory, C., Lough, S., Stone, V., Erzinclioglu, S., Martin, L., Baron-Cohen, S., & Hodges, J. R. (2002). Theory of mind in patients with frontal variant frontotemporal dementia and Alzheimer's disease: theoretical and practical implications. *Brain, 125*(4), 752-764.

Grimm, H. (2000). *Sprachentwicklungstest für zweijährige Kinder (SETK-2). Diagnose rezeptiver und produktiver Sprachverarbeitungsfähigkeiten.* Hogrefe.

Grosse Wiesmann, C., Friederici, A. D., Singer, T., & Steinbeis, N. (2017). Implicit and explicit false belief development in preschool children. *Developmental Science, 20*(5), e12445.

Grosso, S. S., Schuwerk, T., Kaltefleiter, L. J., & Sodian, B. (2019). 33-month-old children succeed in a false belief task with reduced processing demands: A replication of Setoh et al.(2016). *Infant Behavior and Development, 54*, 151-155.

Hacquard, V., & Lidz, J. (2019). Children's attitude problems: Bootstrapping verb meaning from syntax and pragmatics. *Mind & Language, 34*(1), 73-96.

Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science, 304*(5669), 438-441.

Handy, T. C. (Ed.). (2005). *Event-related potentials: A methods handbook.* MIT press.

Harris, P. L. (2006). Social cognition. In D. Kuhn & R. Siegler (Eds.), *Handbook of Child Psychology* (5th ed., pp. 811–858). New York: Wiley.

Harris, P. L., Bartz, D. T., & Rowe, M. L. (2017). Young children communicate their ignorance and ask questions. *Proceedings of the National Academy of Sciences, 114*(30), 7884-7891. https://doi.org/10.1073/pnas.1620745114

Harris, P. L., de Rosnay, M., & Pons, F. (2005). Language and children's understanding of mental states. *Current Directions in Psychological Science, 14*(2), 69-73.

Hirao, K., Miyata, J., Fujiwara, H., Yamada, M., Namiki, C., Shimizu, M., ... & Murai, T. (2008). Theory of mind and frontal lobe pathology in schizophrenia: a voxel-based morphometry study. *Schizophrenia Research, 105*(1-3), 165-174.

Houston-Price, C., & Nakai, S. (2004). Distinguishing novelty and familiarity effects in infant preference procedures. *Infant and Child Development: An International Journal of Research and Practice, 13*(4), 341-348.

Hughes, C., & Ensor, R. (2007). Executive function and theory of mind: Predictive relations from ages 2 to 4. *Developmental Psychology, 43*(6), 1447.

Jiang, Q., Wang, Q., Li, P., & Li, H. (2016). The neural correlates underlying belief reasoning for self and for others: evidence from ERPs. *Frontiers in Psychology, 7*, 1501.

Johnson, C. N. (1982). Acquisition of mental verbs and the concept of mind. L*anguage Development, 1*, 445-478.

Jolliffe, D., & Farrington, D. P. (2004). Empathy and offending: A systematic review and meta-analysis. *Aggression and Violent Behavior, 9*(5), 441-476.

Kaltefleiter, L. J., Sodian, B., Schuwerk, T., & Kristen-Antonow, S. (2022). "I know what's inside because you showed me": Training 33-month-old children to attribute knowledge and ignorance to themselves and others. *Infant Behavior and Development, 68*, 101744.

Kidd, E., Lieven, E. V., & Tomasello, M. (2010). Lexical frequency and exemplar-based learning effects in language acquisition: Evidence from sentential complements. *Language Sciences, 32*(1), 132-142.

Knudsen, B., & Liszkowski, U. (2012). 18-month-olds predict specific action mistakes through attribution of false belief, not ignorance, and intervene accordingly. *Infancy, 17*(6), 672-691.

Kobayashi, C., Glover, G. H., & Temple, E. (2007). Children's and adults' neural bases of verbal and nonverbal 'theory of mind'. *Neuropsychologia, 45*(7), 1522-1532.

Koornneef, A. W., & Van Berkum, J. J. A. (2006). On the use of verb-based implicit causality in sentence comprehension: Evidence from self-paced reading and eye tracking. *Journal of Memory and Language, 54*(4), 445–465. https://doi.org/10.1016/j.jml.2005.12.003

Kormi-Nouri, R. E. Z. A., Nilsson, L.G., & Ohta, N. (2005). The novelty effect: Support for the novelty-encoding hypothesis. *Scandinavian Journal of Psychology, 46*(2), 133-143.

Kovács, Á. M. (2009). Early bilingualism enhances mechanisms of false-belief reasoning. *Developmental Science, 12*(1), 48-54.

Kristen, S., Sodian, B., Licata, M., Thoermer, C., & Poulin-Dubois, D. (2012). The development of internal state language during the third year of life: A longitudinal parent report study. *Infant and Child Development, 21*(6), 634-645.

Kristen-Antonow, S., Jarvers, I., & Sodian, B. (2019). Preschoolers' developing understanding of factivity in mental verb comprehension and its relation to first- and second-order false belief understanding: A longitudinal study. *Journal of Cognition and Development, 20*(3), 354–369. https://doi.org/10.1080/15248372.2019.1586710

Kühn-Popp, N., Sodian, B., Sommer, M., Döhnel, K., & Meinhardt, J. (2013). Same or different? ERP correlates of pretense and false belief reasoning in children. *Neuroscience, 248*, 488-498.

Kulke, L., & Rakoczy, H. (2018). Implicit Theory of Mind–An overview of current replications and non-replications. *Data in Brief, 16*, 101-104.

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.

Lang, M., Lang, W., Uhl, F., Kornhuber, A., Deecke, L., & Kornhuber, H. H. (1987). Slow negative potential shifts indicating verbal cognitive learning in a concept formation task. *Human Neurobiology.*

Leslie, A. M., Friedman, O., & German, T. P. (2004). Core mechanisms in 'theory of mind'. *Trends in Cognitive Sciences, 8*(12), 528-533.

Leslie, A. M. (2005). Developmental parallels in understanding minds and bodies. *Trends in Cognitive Sciences, 9*(10), 459-462.

Levine, M. D. (2009). Differences in learning and neurodevelopmental function in school-age children. In *Developmental-Behavioral Pediatrics* (pp. 535-546). WB Saunders.

Liszkowski, U., Carpenter, M., & Tomasello, M. (2008). Twelve-month-olds communicate helpfully and appropriately for knowledgeable and ignorant partners. *Cognition, 108*(3), 732-739. https://doi.org/10.1016/j.cognition.2008.06.013

Liu, D., Meltzoff, A. N., & Wellman, H. M. (2009a). Neural correlates of belief-and desire-reasoning. *Child Development, 80*(4), 1163-1171.

Liu, D., Sabbagh, M. A., Gehring, W. J., & Wellman, H. M. (2004). Decoupling beliefs from reality in the brain: an ERP study of theory of mind. *Neuroreport, 15*(6), 991-995.

Liu, D., Sabbagh, M. A., Gehring, W. J., & Wellman, H. M. (2009b). Neural correlates of children's theory of mind development. *Child Development, 80*(2), 318-326.

Low, J. (2010). Preschoolers' implicit and explicit false-belief understanding: Relations with complex syntactical mastery. *Child Development, 81*(2), 597-615.

Luck, S. J. (2014). *An introduction to the event-related potential technique.* MIT press.

Luck, S. J., & Gaspelin, N. (2017). How to get statistically significant effects in any ERP experiment (and why you shouldn't). *Psychophysiology, 54*(1), 146-157.

MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk.* Third Edition. Mahwah, NJ: Lawrence Erlbaum Associates.

Mahy, C. E., Moses, L. J., & Pfeifer, J. H. (2014). How and where: Theory-of-mind in the brain. *Developmental Cognitive Neuroscience, 9*, 68-81.

Mani, N., & Plunkett, K. (2007). Phonological specificity of vowels and consonants in early lexical representations. *Journal of Memory and Language, 57*(2), 252-272.

Mani, N., & Plunkett, K. (2008). Fourteen-month-olds pay attention to vowels in novel words. *Developmental Science, 11*(1), 53-59.

Maoz, H., Gvirts, H. Z., Sheffer, M., & Bloch, Y. (2019). Theory of mind and empathy in children with ADHD. *Journal of Attention Disorders, 23*(11), 1331-1338.

Mar, R. A. (2011). The neural bases of social cognition and story comprehension. *Annual Review of Psychology, 62*, 103-134.

Marino, C., & Gervain, J. (2019). The novelty effect as a predictor of language outcome. *Frontiers in Psychology, 10*, 258.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG-and MEG-data. *Journal of Neuroscience Methods, 164*(1), 177-190.

Marno, H., Guellai, B., Vidal, Y., Franzoi, J., Nespor, M., & Mehler, J. (2016). Infants' selectively pay attention to the information they receive from a native speaker of their language. *Frontiers in Psychology, 7*, 1150.

Matthews, D. (2014). *Pragmatic development in first language acquisition.* John Benjamins Publishing Company.

Maxwell, S. E., Lau, M. Y., & Howard, G. S. (2015). Is psychology suffering from a replication crisis? What does "failure to replicate" really mean?. *American Psychologist, 70*(6), 487.

Mayer, R. E., & Moreno, R. (1998). A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of Educational Psychology, 90*(2), 312.

Mecklinger, A., & Pfeifer, E. (1996). Event-related potentials reveal topographical and temporal distinct neuronal activation patterns for spatial and object working memory. *Cognitive Brain Research, 4*(3), 211-224.

Meinhardt, J., Kühn-Popp, N., Sommer, M., & Sodian, B. (2012). Distinct neural correlates underlying pretense and false belief reasoning: Evidence from ERPs. *NeuroImage, 63*(2), 623-631.

Meinhardt, J., Sodian, B., Thoermer, C., Döhnel, K., & Sommer, M. (2011). True-and false-belief reasoning in children and adults: An event-related potential study of theory of mind. *Developmental Cognitive Neuroscience, 1*(1), 67-76.

Meyer, M., Lamers, D., Kayhan, E., Hunnius, S., & Oostenveld, R. (2021). Enhancing reproducibility in developmental EEG research: BIDS, cluster-based permutation tests, and effect sizes. *Developmental Cognitive Neuroscience, 52*, 101036.

Michel, C. M., & Brunet, D. (2019). EEG source imaging: a practical review of the analysis steps. *Frontiers in Neurology, 10*, 325.

Milligan, K., Astington, J. W., & Dack, L. A. (2007). Language and theory of mind: Meta-analysis of the relation between language ability and false-belief understanding. *Child Development, 78*(2), 622-646.

Mirman, D. (2015). psy811: Tools for PSY811 Multilevel Regression. R package version 1.0. http://github.com/dmirman/psy811

Mirman, D. (2017). *Growth curve analysis and visualization using R.* Chapman and Hall/CRC.

Mitchell, J. P., Mason, M. F., Macrae, C. N., & Banaji, M. R. (2006). Thinking about others: The neural substrates of social cognition. *Social Neuroscience: People Thinking About People*, 63-82.

Montag, C., Ehrlich, A., Neuhaus, K., Dziobek, I., Heekeren, H. R., Heinz, A., & Gallinat, J. (2010). Theory of mind impairments in euthymic bipolar patients. *Journal of Affective Disorders, 123*(1-3), 264-269.

Montgomery, D. E. (1992). Young children's theory of knowing: The development of a folk epistemology. *Developmental Review, 12*(4), 410–430. https://doi.org/10.1016/0273-2297(92)90016-U

Moore, C., Bryant, D., & Furrow, D. (1989). Mental terms and the development of certainty. *Child Development, 60*(1), 167-171. https://doi.org/10.2307/1131082

Moriguchi, Y., Ohnishi, T., Mori, T., Matsuda, H., & Komaki, G. E. N. (2007). Changes of brain activity in the neural substrates for theory of mind during childhood and adolescence. *Psychiatry and Clinical Neurosciences, 61*(4), 355-363.

Müller, B. C., Tsalas, N. R., van Schie, H. T., Meinhardt, J., Proust, J., Sodian, B., & Paulus, M. (2016). Neural correlates of judgments of learning - An ERP study on metacognition. *Brain Research, 1652*, 170-177.

Nakano, T., & Kitazawa, S. (2017). Development of long-term event memory in preverbal infants: an eye-tracking study. *Scientific Reports, 7*(1), 1-7.

Németh, N., Mátrai, P., Hegyi, P., Czéh, B., Czopf, L., Hussain, A., ... & Simon, M. (2018). Theory of mind disturbances in borderline personality disorder: A meta-analysis. *Psychiatry Research, 270*, 143-153.

Nuyts, J. (2001). *Epistemic modality, language, and conceptualization: A cognitive-pragmatic perspective.* John Benjamins Publishing.

O'Neill, D. K. (1996). Two-year-old children's sensitivity to a parent's knowledge state when making requests. *Child Development, 67*(2), 659-677.

Oakes, L. M., Baumgartner, H. A., Barrett, F. S., Messenger, I. M., & Luck, S. J. (2013). Developmental changes in visual short-term memory in infancy: Evidence from eye-tracking. *Frontiers in Psychology, 4*, 697.

Oakes, L. M., Hurley, K. B., Ross-Sheehy, S., & Luck, S. J. (2011). Developmental changes in infants' visual short-term memory for location. *Cognition, 118*(3), 293-305.

Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs?. *Science, 308*(5719), 255-258.

Perner, J. (1991). *Understanding the representational mind.* The MIT Press.

Perner, J., & Lang, B. (1999). Development of theory of mind and executive control. *Trends in Cognitive Sciences, 3*(9), 337-344.

Perner, J., & Roessler, J. (2012). From infants' to children's appreciation of belief.

Trends in Cognitive Sciences, 16(10), 519-525.

Perner, J., & Ruffman, T. (2005). Infants' insight into the mind: How deep?. *Science, 308*(5719), 214-216.

Perner, J., Baker, S., Hutton, D., 1994. Prelief: the conceptual origins of belief and pretence. In: Lewis, C., Mitchell, P. (Eds.), *Children's Early Understanding of Mind: Origins and Development.* (pp. 261–286). Lawrence Erlbaum Associates, Inc.

Pernet, C. R., Latinus, M., Nichols, T. E., & Rousselet, G. A. (2015). Cluster-based computational methods for mass univariate analyses of event-related brain potentials/fields: A simulation study. *Journal of Neuroscience Methods, 250*, 85-93.

Pineda-Alhucema, W., Aristizabal, E., Escudero-Cabarcas, J., Acosta-Lopez, J. E., & Vélez, J. I. (2018). Executive function and theory of mind in children with ADHD: A systematic review. *Neuropsychology Review, 28*(3), 341-358.

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clinical Neurophysiology, 118*(10), 2128-2148.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind?. *Behavioral and Brain Sciences, 1*(4), 515-526.

Rakoczy, H. (2012). Do infants have a theory of mind?. *British Journal of Developmental Psychology, 30*(1), 59-74.

R Core Team (2020). *R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.*

Roby, E., & Scott, R. M. (2018). The relationship between parental mental-state language and 2.5-year-olds' performance on a nontraditional false-belief task. *Cognition, 180*, 10-23.

Roehm, D., Schlesewsky, M., Bornkessel, I., Frisch, S., & Haider, H. (2004).

Fractionating language comprehension via frequency characteristics of the human EEG. *Neuroreport, 15*(3), 409-412.

Rose, S. A., Feldman, J. F., & Jankowski, J. J. (2004). Dimensions of cognition in infancy. *Intelligence, 32*(3), 245-262.

Rose, S. A., Feldman, J. F., Jankowski, J. J., & Caro, D. M. (2002). A longitudinal study of visual expectation and reaction time in the first year of life. *Child Development, 73*(1), 47-61.

Rubio-Fernández, P., & Geurts, B. (2013). How to pass the false-belief task before your fourth birthday. *Psychological Science, 24*(1), 27-33.

Ruffman, T., & Perner, J. (2005). Do infants really understand false belief?. *Cognitive Development, 9*, 377-395.

Ruffman, T., Slade, L., & Crowe, E. (2002). The relation between children's and mothers' mental state language and theory-of-mind understanding. *Child Development, 73*(3), 734-751.

Sabbagh, M. A., & Taylor, M. (2000). Neural correlates of theory-of-mind reasoning: an event-related potential study. *Psychological Science, 11*(1), 46-50.

Sabbagh, M. A., Bowman, L. C., Evraire, L. E. & Ito, J. M. B. Neurodevelopmental correlates of theory of mind in preschool children. *Child Development, 8*0, 1147–1162 (2009).

Sabbagh, M. A., Hopkins, S. F., Benson, J. E., & Flanagan, J. R. (2010). Conceptual change and preschoolers' theory of mind: Evidence from load–force adaptation. *Neural Networks, 23*(8-9), 1043-1050.

Sabbagh, M. A., Moulson, M. C., & Harkness, K. L. (2004). Neural correlates of mental state decoding in human adults: an event-related potential study. *Journal of Cognitive Neuroscience, 16*(3), 415-426.

Salvucci, D. D., & Goldberg, J. H. (2000, November). Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications* (pp. 71-78). ACM.

San Juan, V., & Astington, J. W. (2012). Bridging the gap between implicit and explicit understanding: How language development promotes the processing and representation of false belief. *British Journal of Developmental Psychology, 30*(1), 105–122. https://doi.org/10.1111/j.2044-835X.2011.02051.x

Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people: the role of the temporo-parietal junction in "theory of mind". *Neuroimage, 19*(4), 1835-1842.

Saxe, R., Carey, S., & Kanwisher, N. (2004). Understanding other minds. *Annual Review of Psychology, 55*, 87-124.

Saxe, R., & Powell, L. J. (2006). It's the thought that counts: specific brain regions for one component of theory of mind. *Psychological Science, 17*(8), 692-699.

Schüler, A., Scheiter, K., & van Genuchten, E. (2011). The role of working memory in multimedia instruction: Is working memory working during learning from text and pictures?. *Educational Psychology Review, 23*(3), 389-411.

Schurz, M., Radua, J., Aichhorn, M., Richlan, F. & Perner, J. Fractionating theory of mind: a meta-analysis of functional brain imaging studies. *Neuroscience & Biobehavioral Reviews, 42*, 9–34 (2014).

Schuwerk, T., Priewasser, B., Sodian, B., & Perner, J. (2018). The robustness and generalizability of findings on spontaneous false belief sensitivity: A replication attempt. *Royal Society Open Science, 5*(5), 172273.

Scott, R. M., & Baillargeon, R. (2009). Which penguin is this? Attributing false beliefs about object identity at 18 months. *Child Development, 80*(4), 1172-1196.

Scott, R. M., & Baillargeon, R. (2014). How fresh a look? A reply to Heyes. *Developmental Science, 17*(5), 660-664.

Scott, R. M., & Baillargeon, R. (2017). Early false-belief understanding. *Trends in*

Cognitive Sciences, 21(4), 237-249.

Scott, R. M., He, Z., Baillargeon, R., & Cummins, D. (2012). False-belief understanding in 2.5-year-olds: Evidence from two novel verbal spontaneous-response tasks. *Developmental Science, 15*(2), 181-193.

Setoh, P., Scott, R. M., & Baillargeon, R. (2016). Two-and-a-half-year-olds succeed at a traditional false-belief task with reduced processing demands. *Proceedings of the National Academy of Sciences, 113*(47), 13360-13365.

Shrout, P. E., & Rodgers, J. L. (2018). Psychology, science, and knowledge construction: Broadening perspectives from the replication crisis. *Annual Review of Psychology, 69*(1), 487-510.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2016). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. In A. E. Kazdin (Ed.), *Methodological issues and strategies in clinical research* (pp. 547–555). American Psychological Association. https://doi.org/10.1037/14805-033

Sjølie, C., Meyn, E. K., Raudeberg, R., Andreassen, O. A., & Vaskinn, A. (2020). Nonsocial cognitive underpinnings of theory of mind in schizophrenia. *Psychiatry Research, 289*, 113055.

Skoruppa, K., Mani, N., & Peperkamp, S. (2013). Toddlers' processing of phonological alternations: Early compensation for assimilation in English and French. *Child Development, 84*(1), 313-330.

Sodian, B. (2011). Theory of mind in infancy. *Child Development Perspectives, 5*(1), 39-43.

Sommer, M., Döhnel, K., Sodian, B., Meinhardt, J., Thoermer, C., & Hajak, G. (2007). Neural correlates of true and false belief reasoning. *Neuroimage, 35*(3), 1378-1384.

Sommer, M., Meinhardt, J., Eichenmüller, K., Sodian, B., Döhnel, K., & Hajak, G.

(2010). Modulation of the cortical false belief network during development. *Brain Research, 1354*, 123-131.

Southgate, V., Senju, A., & Csibra, G. (2007). Action anticipation through attribution of false belief by 2-year-olds. *Psychological Science, 18*(7), 587-592.

Sprong, M., Schothorst, P., Vos, E., Hox, J., & Van Engeland, H. (2007). Theory of mind in schizophrenia: meta-analysis. *The British Journal of Psychiatry, 191*(1), 5-13.

Surian, L., Caldi, S., & Sperber, D. (2007). Attribution of beliefs by 13-month-old infants. *Psychological Science, 18*(7), 580-586.

Swingley, D., Pinto, J. P., & Fernald, A. (1999). Continuous processing in word recognition at 24 months. *Cognition, 71*(2), 73-108.

Thibaudeau, É., Cellard, C., Reeder, C., Wykes, T., Ivers, H., Maziade, M., ... & Achim, A. M. (2017). Improving theory of mind in schizophrenia by targeting cognition and metacognition with computerized cognitive remediation: A multiple case study. *Schizophrenia Research and Treatment, 2017.*

Tomasello, M., & Haberl, K. (2003). Understanding attention: 12-and 18-month-olds know what is new for other persons. *Developmental Psychology, 39*(5), 906.

Van Overwalle, F. (2009). Social cognition and the brain: a meta-analysis. *Human Brain Mapping, 30*(3), 829-858.

Vanwoerden, S., Greiner, I., Ensink, K., & Sharp, C. (2019). The relations between self-and caregiver-focused reflective function and theory of mind in the context of borderline pathology in adolescence. *Psychiatry Research, 273*, 274-280.

Vogeley, K., Bussfeld, P., Newen, A., Herrmann, S., Happé, F., Falkai, P., ... & Zilles, K. (2001). Mind reading: neural mechanisms of theory of mind and self-perspective. *Neuroimage, 14*(1), 170-181.

Von Holzen, K., & Mani, N. (2012). Language nonselective lexical access in bilingual toddlers. *Journal of Experimental Child Psychology, 113*(4), 569-586.

Wang, Y. W., Lin, C. D., Yuan, B., Huang, L., Zhang, W. X., & Shen, D. L. (2010). Person perception precedes theory of mind: an event related potential analysis. *Neuroscience, 170*(1), 238-246.

Wellman, H. M. (2002). Understanding the psychological world: Developing a theory of mind. In U. Goswami (Ed.), *Handbook of childhood cognitive development* (pp. 167–187). Oxford, UK: Blackwell.

Wellman, H. M. (2011). Developing a theory of mind. In U. Goswami (Ed.), *The Wiley-Blackwell handbook of childhood cognitive development* (pp. 258–284). Wiley-Blackwell.

Wellman, H. M. (2014). *Making minds: How theory of mind develops.* Oxford University Press.

Wellman, H. M., & Bartsch, K. (1988). Young children's reasoning about beliefs. *Cognition, 30*(3), 239-277.

Wellman, H. M., & Cross, D. (2001). Theory of mind and conceptual change. *Child Development, 72*(3), 702-707.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: the truth about false belief. *Child Development, 72*(3), 655-684.

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.

Yalcin, S. (2007). Epistemic Modals. *Mind, 116*(464), 983–1026.

Zhang, T., Sha, W., Zheng, X., Ouyang, H., & Li, H. (2009). Inhibiting one's own knowledge in false belief reasoning: an ERP study. *Neuroscience Letters, 467*(3), 194-198.

Zhang, T., Zhang, Q., Li, Y., Long, C., & Li, H. (2013). Belief and sign, true and false: the unique of false belief reasoning. *Experimental Brain Research, 231*(1), 27-36.

Acknowledgments

I would like to express my gratitude to Prof. Paul Sauseng, for accepting this role and providing positive feedback and motivation to guide me towards the finalization of this work. A special thank you goes to Dr. Tobias Schuwerk for challenging classic paradigms of Theory of Mind and offering his expertise through the whole Crossing project. I would also like to show my deep appreciation to Prof. Nivedita Mani for the long hours spent together analysing eye-tracking data.

I wish to acknowledge the help and incredible work provided by the technical staff in the Developmental Psychology Department of the LMU University through the years. A particular thank you goes to Dr. Victoria Licht for her invaluable assistance in the implementation of tasks and execution of all experiments presented here and to Dr. Carolina Pletti for her motivational boost and precious feedback on EEG methodology. Thank you also to all our patient participants, young and old.

I am so grateful to the DTP coordination, in particular to Prof. Moritz Heene, Prof. Markus Paulus, and Dr. Carolin Schwab, for believing in me: without your support this dissertation would have not been possible!

At last, I would also like to thank my friends and family, who were there for me through the years and were vital with their encouraging words to achieve this goal.