Developmental Origins of Theory of Mind in the Brain: Neural Correlates of False Belief Understanding from Infancy to Early Childhood

Dissertation

zum Erwerb des Doctor of Philosophy (Ph.D.) am

Munich Center of the Learning Sciences

der Ludwig-Maximilians-Universität München

vorgelegt von

Shuting Li

München 2025

1st Supervisor / Erstgutachterin: Prof. Dr. Beate Sodian

2nd Supervisor / Zweitgutachter: PD Dr. Jörg Meinhardt

3rd Supervisor / Weiterere Betreuerin: Dr. Barbara C. N. Müller

Tag der mündlichen Prüfung: February 21, 2025

Acknowledgments

The research presented in this dissertation was supported by the Deutsche Forschungsgemeinschaft (DFG), and my scholarship was funded by the China Scholarship Council (CSC). I would like to extend my sincere gratitude to both institutions for providing invaluable opportunities throughout my doctoral journey. I am also grateful to the Erasmus+ “Short-Term Doctoral Mobility” program, which allowed me a travel bursary during my Ph.D. research.

This dissertation would not have been completed without the tremendous support of many people. First and foremost, I would like to express my deepest appreciation to my supervisor, Prof. Dr. Beate Sodian. Since the start of my doctoral journey, you have provided unwavering support and encouragement, sharing your expertise to help me. I feel incredible lucky to be your student. Your rigorous attitude toward research, along with your selfless commitment to supporting students, has been a role model for me to follow throughout my doctoral career and beyond.

I would also like to extend my heartfelt gratitude to my supervisor, PD Dr. Jörg Meinhardt. Thank you for teaching me electrophysiological methods and for the invaluable knowledge I have gained through our discussions. You have always been generous in sharing your insights and expertise, and you have introduced me to numerous opportunities that have helped me improve my ERP technique skills.

I also extend my heartfelt thanks to my supervisor, Dr. Barbara C. N. Müller. Thank you for meeting with me regularly to discuss my research progress. Every conversation with you has fostered my passion for academic research. I am profoundly thankful for your guidance, both academically and personally, as well as for the opportunity to visit Radboud University Nijmegen.

Special thanks are also extended to Prof. Dr. Markus Paulus for offering me to attend group research colloquia, where I had the opportunity to learn about others' research and share my own. I deeply appreciate your insightful comments and thought-provoking questions.

I am also immensely grateful to my friends, colleagues, and the faculty and staff of the department for making my time at Ludwig-Maximilians-Universität München an unforgettable experience. I cherish the moments I shared with Dr. Özgün Köksal, and I value our conversations as well as the emotional support we provided to each other throughout our research journeys. Thank you for comforting and encouraging me and always be there to support me. No matter where life takes me after LMU, I will always cherish these moments and remember these wonderful people, as this has been one of the most rewarding periods of my life.

I am forever thankful to my parents and my brother for their firm support and belief in me. Though my parents did not receive much formal education, they recognize the value of education and support me in studying abroad and pursuing my ambitions. My family's encouragement has been a constant source of strength.

Lastly, I want to express my deepest gratitude to my boyfriend, Zhidong Lu. You are not only a partner in academic research but also the love of my life. You have stood by me through all the ups and downs, and your support has helped me navigate difficult times. Throughout this challenging yet fulfilling Ph.D. journey, I am grateful to have you by my side, as we support each other in ways that make all the difference.

Elaborate Summary

An essential component of human social interaction and communication is the ability to comprehend mental states, encapsulated in the concept of "Theory of Mind" (Harman, 1978; Perner, 1991). Theory of Mind refers to understanding mental states such as beliefs, desires, and intentions and their influence on behavior (Perner, 1991; Wellman et al., 1990). Over the past 40 years, the exploration of how children develop a Theory of Mind has been a central focus in developmental psychology (Rakoczy, 2022; Sodian et al., 2020). A key aspect of Theory of Mind that has received extensive attention is the false belief understanding—the understanding that others may hold beliefs about the world that differ from one's own knowledge of reality.

Seminal empirical research and a comprehensive meta-analysis demonstrated that children typically begin to attribute false beliefs to themselves and others in the fourth years of life (Wellman et al., 2001; Wimmer & Perner, 1983) when they start to acquire a proper notion of mental representation (Perner, 1991). These traditional findings of a marked age-related trend support the view that conceptual changes in false belief understanding occur at approximately 4 years of age (e.g., Wellman et al., 2001). However, the claim that the false belief understanding undergoes a dramatic conceptual change at this age has been contested (Baillargeon et al., 2010). Behavioral evidence challenging the conceptual change perspective has been demonstrated in both implicit and low-demands explicit false belief tasks. Implicit false belief tasks involve inferring children's understanding of an agent's false belief from their spontaneous behaviors as they observe scenes unfold (Baillargeon et al., 2010; Onishi & Baillargeon, 2005), while explicit false belief tasks require participants to provide direct responses to questions about an agent's false belief (Baron-Cohen et al., 1985; Wellman et al., 2001; Wimmer & Perner, 1983). For instance, nonverbal spontaneous-response paradigms revealed that infants (under 2 years old) and toddlers (ages 2–3 years) exhibit implicit reasoning about false beliefs (e.g., Clements & Perner, 1994; Schneider et al., 2012; Scott & Baillargeon, 2017; Southgate et al., 2007). Moreover, recent research utilizing a behaviorally low-demands false belief task suggests that toddlers aged 2.5 years can grasp explicit false beliefs when task processing demands related to response generation and inhibitory control are reduced properly (Grosso et al., 2019; Setoh et al., 2016; Sodian et al., 2024). If infants and toddlers already exhibit sensitivity to others' false beliefs, it is plausible that the cortical regions involved in false belief understanding are engaged in these socio-cognitive functions, not only during toddlerhood but potentially as early as infancy.

Earlier approaches to exploring false belief understanding were primarily focused on the behavioral level, whereas recent approaches have further investigated its neuroscientific basis. Over the last two decades, significant progress has been made toward identifying a specialized neural system associated with false belief understanding, primarily focusing on adults and children at 6 to 12 years of age (e.g., Gweon et al., 2012; Meinhardt et al., 2011; Sommer et al., 2010). Neuroimaging studies revealed that the brain regions implicated in false belief understanding are primarily located in the frontal regions (e.g., Geangu et al., 2013; Sommer et al., 2010) and the posterior regions, often measured in parietal or parieto-occipital regions (e.g., Grosse Wiesmann et al., 2020; Perner et al., 2006).

Researchers have begun exploring the neural correlates of this socio-cognitive ability in children under 6 to address important open questions in the development of false belief understanding. Although such studies remain insufficient, they provide important insights (e.g., Grosse Wiesmann et al., 2020; Richardson et al., 2018; Richardson & Saxe, 2020). For example, Richardson et al. (2018) and Moraczewski et al. (2018) observed that in children aged 3 to 6, the temporal dynamics of the mentalizing network closely mirrored those seen in adults, contrasting with other control networks. Grosse Wiesmann et al.

(2017b) reported increased connectivity in tracts surrounding mentalizing regions, including the medial prefrontal cortex (MPFC) and the temporoparietal junction (TPJ), in 3- and 4-year-olds. Furthermore, Hyde et al. (2018) demonstrated that the right TPJ is preferentially active when children are thinking about others' thoughts, even in infants as young as 7 months.

Existing magnetic resonance imaging (MRI) studies have aimed to "localize" neural regions involved in false belief reasoning with high spatial resolution. However, due to the limited temporal resolution of MRI techniques, these studies offer only a partial understanding of how these regions contribute to the cognitive processes underlying false belief understanding. To gain further insights into the temporal dynamics of these processes, electroencephalogram (EEG) techniques, which provide higher temporal precision, may offer a complementary approach. A notable event-related potential (ERP) study has examined belief reasoning in children around 4 years of age, a developmental milestone associated with the emergence of traditional explicit false belief understanding (Liu et al., 2009b). Liu et al. (2009b) compared neural responses between children aged 4-6 years and adults during tasks involving belief attribution to story protagonists and reality judgments. Children were grouped into "passers" and "failers" based on an independent assessment of their behavioral false belief competence. A late waveform was observed over frontal regions only for child "passers," with a less localized and more diffuse scalp distribution than in adults. Children who failed the behavioral false belief task showed no systematic differentiation between belief and reality conditions on the neural level. To date, this study provides preliminary evidence of brain-behavior connections in acquiring false belief understanding in early childhood. However, Liu et al. (2009b) did not differentiate between the false and true beliefs, treating them as a combined condition in the ERP belief task. Given that children demonstrated explicit false belief competence in a low-demands false belief task, it is necessary and critical to investigate whether a specialized neural system supports false belief processing in younger children. This dissertation presents a first step in extending the line of Liu et al.’s (2009b) study to children under the age of 3 years; it takes a more direct and rigorous approach to examining the neural basis of false belief understanding.

Furthermore, it is important to acknowledge that task-dependent and task-independent neural techniques differ fundamentally in their approach and should not be equated when interpreting findings across studies. Task-dependent ERP techniques involve recording the time-locked brain activity to a specific external stimulus designed to elicit a false belief judgment (Sabbagh, 2013). In contrast, task-independent techniques, like resting-state EEG, measure brain activity without externally induced stimuli, focusing instead on intrinsic neural processes (MacLean et al., 2012). They highlight the degree of integration processes within the brain (i.e., brain coherence, Aykan et al., 2021) or hemispheric asymmetry in brain activity (e.g., Licata et al., 2015), which are relevant for understanding socio-cognitive and emotional processes. Differences in frontal and parietal asymmetry might indicate varying levels of engagement in ToM-related processes, which are critical for understanding false beliefs (N. A. Fox et al., 1995; Stewart et al., 2011). For example, a resting-state source-localized EEG study provided evidence supporting the right-hemispheric lateralization of brain activity linked to false belief understanding (Sabbagh et al., 2009). Sabbagh et al. (2009) identified individual differences in alpha oscillations in the dorsal MPFC and several right hemisphere areas—including the TPJ, precentral gyrus, cuneus, and inferior temporal cortex—were associated with representational Theory of Mind performance in 4-year-old children. Therefore, beyond examining task-dependent neural correlates of false belief understanding in toddlers, investigating task-independent neural correlates at around 4 years of age or even younger may facilitate the study of intrinsic brain networks without the influence of explicit tasks. This task-independent neural method could offer an additional understanding of how individual differences in resting-state brain activity during early childhood are associated with the subsequent development of false belief understanding (e.g., Sabbagh et al., 2009).

This dissertation employed both task-dependent and task-independent methods to investigate the neural correlates of false belief understanding from infancy to early childhood. This dissertation aims to address two overarching research questions that advance our understanding of false belief reasoning in early childhood: (1) What neural-behavioral connections underpin false belief understanding in toddlers under 3 years of age? (2) How does the specialized neural system associated with false belief understanding emerge? Specifically, does this specialized neural system develop because of behavioral false belief competence, or is this system already functional and active before behavioral competence becomes observable? To address these questions, this thesis conducted three empirical studies among children aged 14 to 52 months, and the results are presented in subsequent chapters (Chapter 2 and Chapter 3).

Study 1 is a cross-sectional investigation into the brain-behavior connections associated with false belief understanding in 33- to 36-month-old toddlers. A multi-trial explicit false belief task was designed to be compatible with the task-dependent ERP methodology. Toddlers were grouped into passers and failers according to their performance on a low-demands behavioral false belief task (see Setoh et al., 2016). This study aims to examine the neural basis of explicit false belief understanding in this age group and explore potential brain-behavior connections associated with it. The findings from Study 1 revealed distinct neural patterns associated with false belief competence in 33- to 36-month-old children. Specifically, children who demonstrated false belief competence exhibited a more bilaterally diffused occipital positive late waveform in the false belief condition compared to their peers without false belief competence. Additionally, a late negative waveform, predominantly over right-lateralized frontocentral sites, consistently differentiated the false belief from the true belief condition, regardless of performance on low-demands behavioral false belief task. These findings provide evidence of neural correlates associated with false belief competence in children under 3 years old. Furthermore, they highlight a developmental pattern linking occipital positive late waveforms to the development of false belief understanding.

Studies 2 and 3 adopted a longitudinal design to investigate the relationship between resting-state EEG alpha asymmetry in infants and toddlers and their false belief understanding at around 3 and 4 years of age. This investigation distinguished between explicit and implicit false belief understanding. Building on evidence suggesting that cortical network supporting false belief understanding may begin to develop as early as 3 years of age (e.g., Grosse Wiesmann et al., 2020; Richardson et al., 2018; Richardson & Saxe, 2020), and the potential right-hemispheric lateralization of brain activity associated with false belief understanding in infants and toddlers (Hyde et al., 2018; Sabbagh et al., 2009), following exploratory research questions are posed: (1) Is there a correlation between brain asymmetric activity observed during resting-state EEG recording and false belief understanding? (2) Does resting-state EEG asymmetry precede the development of representational false belief understanding, and can it predict later behavioral performance in false belief tasks? Employing a longitudinal design, Study 2 assessed resting-state EEG alpha asymmetry across frontal and parietal electrode sites at 34 months, explicit false belief understanding at 52 months, and implicit false belief understanding at both time points. Study 3 analyzed data from another independent longitudinal dataset to test the generality of the relationship between resting-state EEG alpha asymmetry (assessed at 14 months) and explicit false belief understanding at 51 months old. Results from both studies showed that better explicit false belief understanding at age 4 was associated with greater right than left frontal activity at 14 and 34 months. Better implicit false belief understanding was only cross-sectionally associated with greater relative right than left parietal activity at 34 months. These findings suggest resting-state EEG alpha asymmetry may be a stable early-developing neural marker in explicit false belief understanding. Additionally, these findings tentatively suggest that implicit false belief understanding may not follow a monotonic development across childhood and that implicit and explicit false belief understanding may develop based on partly distinct neural mechanisms.

Overall, this dissertation makes novel contributions to identifying the task-dependent and task-independent neural correlates of false belief understanding from infancy to early childhood through electrophysiological techniques. In Study 1, toddlers' behavioral competencies in a low-demands false belief task were significantly correlated with neural responses linked to false belief understanding. This study is the first to investigate task-dependent ERP correlates of false belief understanding in children aged 3 years and younger, raising the possibility that a sensitive neural system supporting false belief understanding may emerge early in development. Studies 2 and 3 further indicate that resting-state EEG alpha asymmetry assessed at infancy and toddlerhood may serve as an early neural precursor to subsequent explicit false belief understanding at 4 years of age—the typical age when explicit false belief understanding emerges. These studies offer preliminary evidence for a neural marker that reliably predicts individual differences in false belief understanding longitudinally. Additionally, comparisons between Study 1 and subsequent Studies 2 and 3 suggest that the brain–behavior connections associated with false belief understanding in preschool-aged children are not merely the result of the short-term acquisition of behavioral competence; Instead, they may reflect the early developmental (ontogenetic) and possibly evolutionary (phylogenetic) emergence of the Theory of Mind network. Collectively, these studies bridge a critical gap between the extensive research on the early behavioral development of false belief understanding and its corresponding neural profiles. Brain regions associated with Theory of Mind exhibit early signs of functional specialization even before children succeed in explicit false belief tasks. However, these findings are exploratory and require further validation, primarily through future research exploring the neural mechanisms underpinning both implicit and explicit false belief understanding in children younger than 4 years of age across multiple developmental stages.

Table of Contents

List of Tables

List of Figures

Abbreviations

ToM	Theory of Mind
DMN	Default Mode Network
EEG	Electroencephalogram
EF	Executive Function
ERP	Event-Related Potential
FB	False Belief
FBU	False Belief Understanding
MPFC	Medial Prefrontal Cortex
MRI	Magnetic Resonance Imaging
MTG	Middle Temporal Gyrus
PCC	Posterior Cingulate Cortex
PET	Positron Emission Tomography
STG	Superior Temporal Gyrus
STS	Superior Temporal Sulcus
TPJ	Temporoparietal Junction

Chapter 1: General Introduction

Effective social interaction relies on Theory of Mind abilities, which involve understanding not only observable behavior but also the underlying mental states, such as beliefs and intentions. Over the past 40 years, extensive behavioral and neuroimaging research has provided substantial evidence on the developmental processes of Theory of Mind abilities. An essential milestone in Theory of Mind development is the ability to understand that others may hold beliefs that do not align with reality (e.g., Wimmer & Perner, 1983). This socio-cognitive ability, known as false belief understanding, involves inferring others' mental states independently of one's own knowledge. Understanding how the developing brain attributes mental states to others and enables us to understand and predict others' behavior is a compelling endeavor. Furthermore, exploring the development of false belief understanding is important for enhancing our understanding of how the human mind works and for fostering false belief reasoning skills in young children. Fostering false belief understanding is the cognitive basis for complex social interactions such as cooperation and competition and children's social circumstances at school, including academic performance and motivation (Wellman, 2018). Additionally, autism spectrum disorder, a biologically based condition characterized by specific impairments in false belief understanding ability (Perner & Leekam, 2008), has also attracted widespread attention (e.g., Baron-Cohen, 2000, 2001). Therefore, comprehending the development and neural basis of false belief understanding is crucial for elucidating complex human social behaviors like cooperation, competition, and pedagogy, as well as providing insights into the neuropathology of developmental disorders such as autism.

Since the 1980s, research in cognitive development has investigated children's ability to understand others' beliefs (see Wimmer & Perner, 1983, for location false belief reasoning; Hogrefe et al., 1986, for content false belief reasoning). Traditionally, it was believed that a developmental breakthrough occurred around 4 years of age when children start explicitly reasoning about others' false beliefs (Wellman et al., 2001; Yu & Wellman, 2024). However, this perspective has been contested by later studies demonstrating implicit false belief understanding in infancy (e.g., Onishi & Baillargeon, 2005) and explicit false belief understanding in toddlers as young as 2.5 years old when processing demands are reduced (e.g., Setoh et al., 2016). Since the 1990s, studies have attempted to investigate the neural correlates of Theory of Mind in adults (e.g., Baron-Cohen et al., 1994; Fletcher et al., 1995; Goel et al., 1995). Although numerous neural studies on Theory of Mind provided converging evidence for a consistent set of brain regions, known as the "mentalizing network," that are involved in various Theory of Mind tasks in adults and older children aged 6-12 years (Molenberghs et al., 2016; Schurz et al., 2014, 2021), investigating the neural underpinnings in children under 6 years of age remains methodologically challenging (Sabbagh, 2013). While some initial studies have begun exploring this area, the evidence concerning very young children is still notably insufficient (Richardson & Saxe, 2020a). Understanding others' ability to hold false beliefs about the world is a critical component of Theory of Mind and serves as a fundamental test for assessing its development in children, so false belief tasks are of particular significance. In light of this, the present dissertation focuses on false belief tasks to explore the neural correlates associated with false belief understanding from infancy to early childhood. This research uses cross-sectional and longitudinal designs to address task-dependent neural correlations (i.e., ERP) and task-independent neural correlations (i.e., resting-state EEG alpha asymmetry).

This dissertation begins with a general introduction focused on Theory of Mind and false belief understanding in Chapter 1. The chapter begins by defining Theory of Mind and offering an overview of theoretical perspectives on its emergence. Subsequently, it presents behavioral evidence of early false belief understanding. It also discusses theoretical perspectives on the relationships between implicit and explicit false belief understanding, supported by longitudinal findings and relations to other cognitive abilities. It then reviews neural studies that illustrate the state-of-the-art neural underpinnings related to the development of false belief understanding. This chapter ends by clarifying the research questions and outlining the research design. Chapters 2 and 3 present three empirical studies that examine task-dependent and task-independent neural correlates of false belief understanding from infancy to early childhood are presented. Each chapter introduces the current state of the literature concerning specific research questions, followed by a description of the methodologies, results, and a discussion of the findings for each empirical study. The general discussion in Chapter 4 summarizes the findings of the three studies and discusses the overarching research questions. It then discusses the potential theoretical implications of the present findings. Finally, it presents the limitations of the research, outlines open questions for future investigation, and concludes with a summary.

1.1. Theory of Mind

1.1.1. The definition of Theory of Mind and brief historical framing

Theory of Mind is also referred to as “mentalizing” or “mindreading.” The concept of Theory of Mind was first articulated by Premack and Woodruff (1978) in their seminal article, “Does the chimpanzee have a Theory of Mind?” These authors argued that “the individual imputes mental states to himself and others. A system of inferences of this kind is properly viewed as a theory, first, because such states are not directly observable, and second because the system can be used to make a prediction, specifically, about the behavior of other organisms” (p. 515). According to this definition, three criteria—borrowed from the philosophy of science—must be met to justify the use of the term “theory.”

A theory must (1) constitute a *coherent* body of knowledge, (2) make an *ontological* commitment about what kinds of things there are in the world, and (3) provide a *causal-explanatory framework* to account for the phenomena within its domain (Perner, 1991; Wellman & Bartsch, 1988). The body of social knowledge is referred to as the Theory of Mind for several reasons (Saracho, 2014). First, mental states are invisible, rendering their existence theoretical (Stich, 1983). As Premack and Woodruff (1978) explain, individuals attribute mental states to others, constructing a theory about minds. Wellman (1990) argues that our understanding of minds assumes the structure of a theory because it makes precise ontological distinctions, provides a causal-explanatory framework, and relates its paradigm to other paradigms within the theoretical framework. However, “Theory of Mind” is often used loosely without implying a formal, structured theory in the sense of a body of knowledge (Sodian & Kristen, 2016).

Originally, Premack and Woodruff (1978) applied the term Theory of Mind to chimpanzees rather than humans. In the pioneering study, the researchers presented an adult chimpanzee with videotaped scenarios involving a human protagonist facing various challenges, along with photographs depicting potential solutions. The chimpanzee’s consistent selection of the correct photographs suggested that it could recognize the situation as problematic, comprehend the agent’s intentions, and accordingly choose solutions that aligned with those intentions. This experiment provided initial evidence that non-human primates could infer mental states, which is a foundational aspect of Theory of Mind. Woodruff and Premack (1979) also investigated tactical deception in chimpanzees, as such behavior is another valid approach to demonstrating the presence of a Theory of Mind. Deceptive actions, while having high adaptive value due to their strategic flexibility, also require the ability to conceptualize the deceived individual’s false belief as part of one’s planning (Sodian, 1991; Sodian et al., 1991; Sodian & Frith, 1992). Three philosophers commented on Premack and Woodruff’s (1978, 1979) study, pointing out that its design did not exclude the possibility that behavioral learning, as opposed to mental state attribution, could be the cause of the chimpanzee’s successful actions (Bennett, 1978; Dennett, 1978; Harman, 1978). They proposed criteria that are comparable to the current standard location false belief task (see Section 1.2.1.1) as a possible paradigm for determining whether a person or a primate actually holds a Theory of Mind: The participant is aware that they, along with the other agent, observe a particular state of affairs X. Subsequently, in the absence of the other agent, the participant witnesses an unexpected change in the state of affairs from X to Y. The participant now knows that Y is the current state but also understands that the other agent still believes X to be the case. Initially, Premack and Woodruff’s (1978) studies did not consider children’s Theory of Mind abilities.

Nevertheless, after 1978, the term Theory of Mind rapidly gained traction among researchers probing into the cognitive underpinnings of children’s understanding of mental states. The first systematic study of children’s comprehension of explicit false beliefs was carried out in 1983 by Wimmer and Perner, and it is well-known as “Max’s Chocolate Story.” This study adhered to theoretical frameworks proposed by Bennett (1978), Dennett (1978), and Harman (1978), leading to the creation of the standard location false belief task, also referred to as the *change-of-location task* or *unexpected-transfer task*. This task was designed to evaluate the existence of Theory of Mind in children, laying a seminal foundation for exploring socio-cognitive competencies.

During the initial phase of research on Theory of Mind, two conferences in 1986 further incubated the field. The volume of papers from these conferences provided an overview of the early developmental Theory of Mind research (Astington et al., 1988). The subsequent publication of influential books by Wellman (1990) and Perner (1991) further propelled researchers' interest in Theory of Mind. Additionally, Wellman et al. (2001) conducted a meta-analysis of over 500 studies on false belief understanding in young children, demonstrating a robust developmental trend despite the facilitating effects of various task manipulations. Specifically, children aged 2.5 years to early 3 years tended to respond based on reality, while a significant increase in correct, belief-based responses emerged after the age of 3.5 years (Wellman et al., 2001). This developmental trend was consistent across different contexts, regardless of whether the test question pertained to mental states or behavior and whether the protagonist was portrayed as a story character, a video actor, a doll, a child, or an adult (Sodian & Kristen, 2016). These findings support the perspective that explicit Theory of Mind emerges through conceptual change processes. This conceptual change can be characterized either as (1) a transition from a simple desire-based psychology to a belief–desire naïve psychology (Wellman, 1990) or (2) a shift from a situation-based to a representation-based understanding of behavior (Perner, 1991). The following section reviews theoretical frameworks that account for the emergence of Theory of Mind.

1.1.2. Theoretical perspectives on the emergence of Theory of Mind

Given the evidence that children begin to understand explicit false beliefs around age 4, various researchers proposed theories to explain the emergence of Theory of Mind within the broader context of socio-cognitive development. This section reviews the theoretical frameworks that aim to explain the developmental milestones observed in standard explicit false belief tasks.

Theory-Theory view

The Theory-Theory perspective on the emergence of Theory of Mind, as proposed by Wellman and colleagues (Gopnik & Wellman, 1992, 2012), conceptualizes children's development of understanding others' mental states as akin to the process of scientific theory formation and refinement. According to this view, between the ages of 2½ and 5 years, children undergo a series of transitions in their understanding of others' minds. Initially, 2-year-olds are adept at expressing their own desires and understanding others' desires and perceptions as direct drives toward objects. Wellman (1990) referred to the early form of Theory of Mind in young children as "desire psychology," as the naïve psychological understanding of 2-year-olds primarily revolves around basic desires. At this stage, children lack an understanding of beliefs, meaning they do not yet recognize that individuals hold internal mental representations-convictions as to how the world is. By age 3, children expand their basic desire-based psychology to include a rudimentary understanding of beliefs. Wellman referred to this expanded understanding as "desire-belief psychology" since it incorporates both beliefs and desires. Belief–desire psychology refers to the ability to represent others' beliefs about the world, their desires for how it should be, and their potential rational actions taken to fulfill those desires based on their beliefs (Wellman, 1990). With desire-belief psychology, children can make more sophisticated predictions about behavior, explaining actions using both belief and desire concepts (Wellman, 1990). Between ages 4 and 5, children develop a representational understanding of the mind, which involves forming representations of others' mental states independent of reality. Thus, the conceptual change underlying the emergence of Theory of Mind can be characterized as a transition from a simple desire-based psychology to a belief–desire psychology (Wellman, 1990).

Meta-representation view

The term "meta-representation" is defined as the *representation of a representation*, or more precisely, the *representation of a representation* ***as*** *a representation* (Perner, 1988). This definition refers to the "ability to represent the representation relationship itself," as discussed in the concept of "recursive meta-representational ability" (Pylyshyn, 1978, p. 593). Meta-representation involves mentally modeling the relationship between referent and sense, effectively linking these two aspects. The referent denotes what the representation stands for, whereas the sense refers to the particular manner in which the referent is represented (Perner, 1991). Through meta-representation, Theory of Mind renders a recursive process, allowing for embedding similar elements within one another. It entails embedding representations in the framework of Theory of Mind by understanding the relationship between a misrepresentation and the actual reality (Perner, 1991; Pylyshyn, 1978).

According to the meta-representation view, a fundamental understanding of representations as representations is the foundation for having the ability to comprehend mental states (representations) (Perner, 1991). By recognizing certain cognitive processes as mental representations, a child gains deeper insights into mental functioning, which can be understood through a meta-representational framework. An example of this is mistaken actions—actions driven by a misconception or false belief about reality. Such instances are challenging to grasp without a meta-representational perspective, as false beliefs involve mental misrepresentations of actual situations. These situations necessitate distinguishing between the referent (the actual situation) and the sense (the perceived situation) since actions are based on this misperception. Thus, understanding the difference between the referent and the sense of mental representations serves as a key indicator of a child's ability to recognize representations as representations, which is foundational to developing a Theory of Mind.

In standard false belief tasks, developmental progression indicates that children begin to understand belief as a representational state of mind around age 4 (Wellman et al., 2001; Wimmer & Perner, 1983). Children under age 4 do not simply face challenges with experimental procedures; they genuinely struggle with understanding beliefs. This difficulty arises from a deficit in meta-representational abilities—they cannot comprehend that something can serve as a representation. Understanding a false belief as a misconception of reality requires the ability to distinguish between the referent (the actual reality) and the sense (the way reality is mentally represented). Around age 4, children develop the meta-representational ability to understand that something can be a representation or misrepresentation. Thus, the conceptual change underlying the emergence of Theory of Mind can be characterized as a shift from a situation-based to a representation-based understanding of behavior (Perner, 1991).

Mental files theory

Perner and colleagues further used mental files theory to comprehensively explain children's developing false belief understanding (Doherty & Perner, 2020; Perner et al., 2015; Perner & Leahy, 2016). According to this theory, when a child perceives an object with varying appearances or receives different verbal labels, they may form two distinct mental representations—termed "mental files"—for the same entity. These coreferential files refer to the same object but encapsulate different perspectives. Such mental files can be used to represent the differing viewpoints individuals may hold (e.g., beliefs) and to capture variations in mental perspective. Ordinarily, having distinct files implies the existence of two separate objects; therefore, to convey that only one object is involved, these coreferential files must be linked.

In the traditional explicit false belief task (Wimmer & Perner, 1983), children are presented with the story of Max, who mistakenly believes his chocolate remains in its original location, unaware that it has been moved to a new location. Children are then asked where Max will look for his chocolate. How does a child encode Max's beliefs under such circumstances? For children older than 4 years, their own representation of the chocolate's location is stored in a "regular file," which reflects the updated, true location. In contrast, a "vicarious file" indexed to Max contains his outdated belief about the chocolate's original location. Vicarious files are coreferentially linked to regular files in order to establish the sameness of referent between linked files (Perner et al., 2015; Perner & Leahy, 2016). On the regular file, one registers everything that happens to the chocolate; on the vicarious file, only events witnessed by Max (Perner et al., 2007; Perner & Roessler, 2012). This theory is illustrated in **Figure 1**, where a 5-year-old child uses regular files to store her own information about the chocolate and a coreferentially linked vicarious file to store Max's information about the chocolate. Once the vicarious mental file includes this information, it can predict what Max might do to retrieve his chocolate, much like I would use my own mental files to determine how to help Max get his chocolate. Based on his vicarious files, Max would go to the original location of the chocolate, whereas I would need to go to the new location to bring it to him. Although children may hold coreferential mental files well before they are able to solve these tasks (typically around age 4), they still lack the ability to effectively link these coreferential files.

Mental files theory is integral not only to children's false belief understanding but also to dual naming, exemplified by the alternative naming game. Around age 4, children begin to pass the false belief task simultaneously as they start to make sense of identity statements (e.g., "the die is the eraser"; Perner et al., 2011). Understanding such identity statements involves processing identity statements, overcoming mutual exclusivity (accepting multiple labels for an object), engaging in visual perspective-taking, and understanding belief differences (Doherty & Perner, 2020).

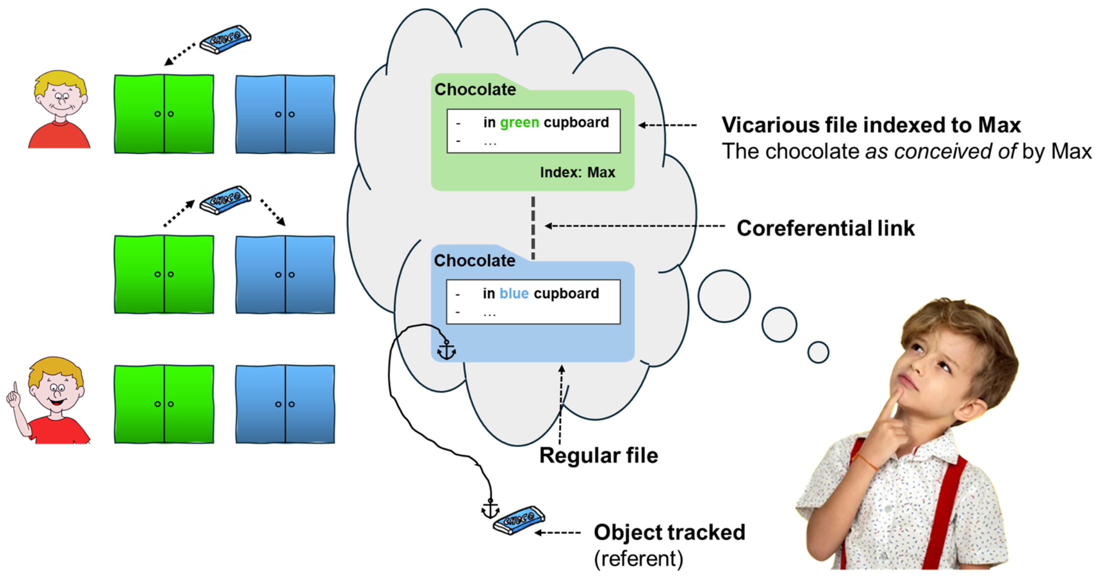


Figure 1. A 5-year-old child uses a vicarious file to track Max's beliefs, as adapted from Perner et al. (2015) and Perner and Leahy (2016). Copyright for the image of the child is held by https://tist.school/blog/top-strategies-boost-childs-imagination-skills.

Modular theory

Modular Theory, also known as the "Mental Module" or "Theory of Mind Module/Mechanism" (ToMM; Leslie, 1987; Leslie et al., 2004), posits that the Theory of Mind reasoning is carried out by one or more specialized neural systems. These systems constitute an innate neural mechanism dedicated to Theory of Mind reasoning, which is present in infancy and gradually 'comes online' throughout development (Carey, 1985; Fodor, 1987; Leslie, 1987). This mental module interprets people's behavior and produces representations of their beliefs and pretenses. Leslie argues that the development of Theory of Mind is not a product of developing concepts or learned simulation abilities but is instead the result of the maturation of the ToMM (Leslie, 1994b, 1994a; Leslie et al., 2004; Scholl & Leslie, 1999, 2001).

Leslie's original formulation emphasizes a critical feature in attributing mental states: decoupling a representation from its reference to the actual state of affairs (U. Frith et al., 1991; Leslie, 1987). As discussed in Leslie's work, the decoupling mechanism, ‘refers to a cognitive process that allows individuals, particularly children, to separate their representations of reality from their false beliefs or pretend representations. This mechanism enables a child to engage in false belief or pretense without confusing it with actual beliefs or knowledge about the world.

Leslie claimed that pretense and false belief understanding are based on similar mechanisms of decoupling mental representation from reality. Pretense understanding constitutes another milestone in Theory of Mind development as it also involves the ability to separate mental (fictional) from actual content. There may be a close psychological relationship between the concepts of belief and pretense, as both might be mediated by the same underlying cognitive mechanism and may originate from a common pre-structured representational system. Pretense may emerge earlier, while the understanding of belief might develop later due to either maturational factors or the need for more complex and less readily accessible information (Leslie, 1988). Furthermore, employing the concepts of pretense and belief may involve distinct performance demands that can only be met at different developmental stages, depending on various influencing factors (Leslie & Thaiss, 1992). However, findings from empirical ERP studies of belief versus pretense reasoning did not support the view that there is a neural similarity between the decoupling of belief from reality and that of pretense from reality (Kühn-Popp et al., 2013; Meinhardt et al., 2012). Nevertheless, the notion of a decoupling mechanism has been cited in the neurocognitive literature, though not necessarily implying an equivalence of pretense and false belief understanding.

1.2. False belief: the way to Theory of Mind

Beliefs, as mental representations influenced by personal experiences, are inherently subjective and can diverge from the objective state of the world, manifesting as either true or false. This divergence occurs when a belief’s representation does not accurately reflect actual reality, termed a “false belief.” A key entailment in Theory of Mind development is the ability to recognize that others may hold false beliefs that differ from the (perceived) state of the world — an inference made independently of one’s own knowledge of reality (Goodman et al., 2006).

The earliest emergence of children’s understanding that others can hold false beliefs has sparked considerable debate about when this ability first appears. According to traditional explicit false belief tasks [1], a significant developmental milestone in Theory of Mind development occurs around age 4, when explicit false belief understanding first appears (Wellman et al., 2001). In contrast, evidence gradually amassed over the past two decades from implicit false belief tasks indicates that false belief understanding may already be present in infants (under age 2 years) and toddlers (age 2–3 years), thereby constituting an integral part of Theory of Mind from an early age (see **Table 1**). This section reviews findings from both explicit and implicit false belief tasks in early childhood, followed by an overview of key theoretical perspectives on the relationship between implicit and explicit false belief understanding, supported by relevant longitudinal empirical evidence and relations to other cognitive abilities.

[1] In the current dissertation, explicit false belief tasks are defined as those that require participants to provide elicited responses to direct questions regarding an agent’s false belief (Baron-Cohen et al., 1985; Wellman et al., 2001; Wimmer & Perner, 1983). In contrast, implicit false belief tasks involve inferring children’s understanding of an agent’s false belief from their spontaneous behaviors as they observe an unfolding scene (Baillargeon et al., 2010; Onishi & Baillargeon, 2005).

Table 1. Summary of the false belief tasks.

Task type	Tasks	Description
Explicit false belief tasks	Location false belief task	Child witnesses an agent store an object in location A, which is subsequently relocated to location B after the agent leaves. The child is then queried about the agent's likely search location for the object.
	Content false belief task	Child sees a distinctive container (e.g., a Smarties box) and looks inside to find what it has (e.g., a pig figurine), then judges what someone who has never seen the container will think it holds.
Implicit false belief tasks	Anticipatory looking task	A child observes an agent acting on beliefs that may be true or false. The child's anticipatory gaze is recorded to determine if it accounts for the agent's mental state in predicting the agent's subsequent actions.
	Violation-of-expectation task	Child observes an agent forming a true or false belief and acting either in line with or contrary to it. The child's looking time is measured and compared between expected and unexpected behavior.
	Prompted action task	Experimenters measure whether participants spontaneously consider the agent's true or false belief in their interactions, either by helping or interpreting the agent's communicative acts.

1.2.1. Behavioral evidence of early false belief understanding

1.2.1.1. Behavioral evidence of explicit false belief understanding

Standard explicit false belief tasks

More than 40 years ago, Wimmer and Perner (1983) introduced the location false belief task, also named the *change-of-location task* and the *unexpected-transfer task*, which became the standard measure for assessing someone's ability to ascribe beliefs to others. Understanding another person's false belief requires explicit representation of the inconsistency of this person's belief about one's own knowledge (Wimmer & Perner, 1983). This ability serves as strong evidence of "mindreading," demonstrating that children can attribute mental states that differ from actual reality. In contrast, understanding true beliefs or desires can be resolved through egocentric assumptions that others share the same knowledge as oneself and may simply rely on observing the actual state of the world and applying mental terms appropriately (Flavell, 2000). In the location false belief task, a target object is moved from one location to another without the knowledge of the story protagonist, who placed the object in its initial location. This task tests the ability of children to understand that the story protagonist has a false belief about the location of the object (Wimmer & Perner, 1983). As shown in **Figure 2** (Perner & Lang, 1999; Wimmer & Perner, 1983), after Max put his chocolate in the green cupboard, he left for the playground; meanwhile, his mother transferred the chocolate from the green cupboard to the blue cupboard while Max was away. Then, his mother went out into the garden. Max came back home for his chocolate. The protagonist (i.e., Max) did not witness the relocation of the target object (i.e., Max’s chocolate) and therefore held a false belief about its location. Participants in a study were asked, “Where will Max look for his chocolate?” There was no consistent evidence for children aged 3 to 4 years, that they could accurately attribute a false belief. Approximately half of the children aged 4 to 6 responded correctly, and almost all children aged 6 to 7 accurately identified the green cupboard as the location where Max would look for the chocolate upon his return (see **Figure 3**). Around the age of 4, children begin to attribute to an agent a subjective misrepresentation—specifically, a false belief that diverges from reality and is inconsistent with their own perspective (Rakoczy, 2022).

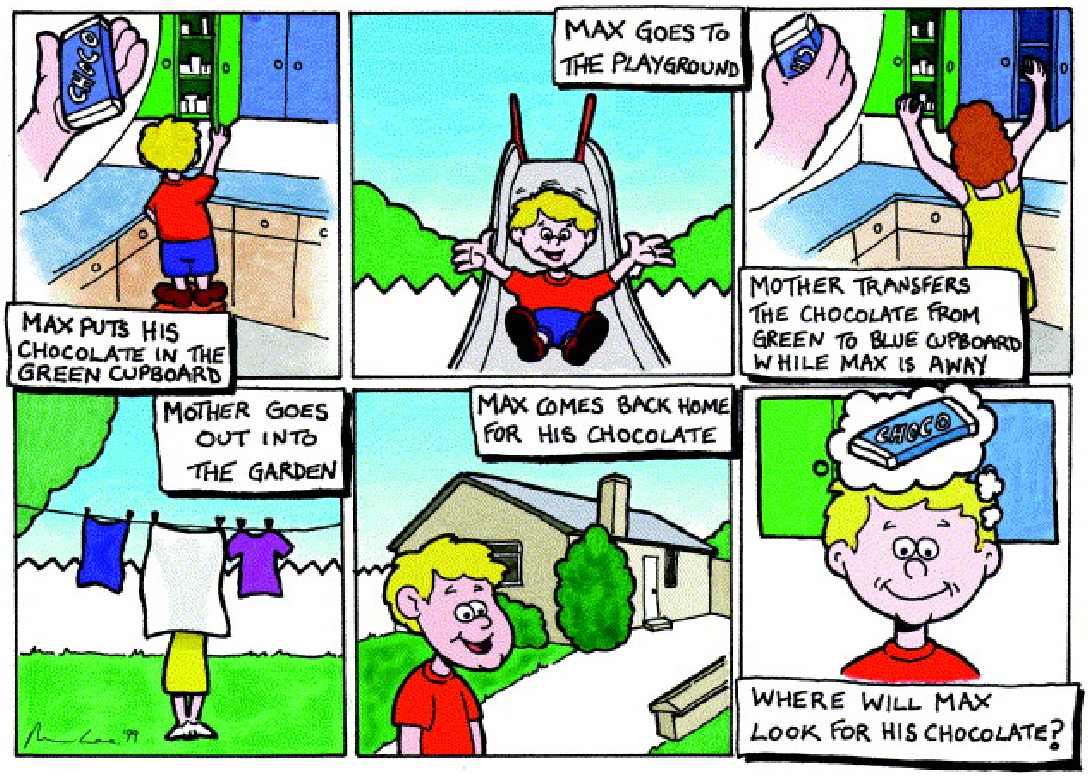


Figure 2. The Max's Chocolate Story exemplifies the location false belief task commonly employed to assess children's Theory of Mind abilities. In this task, children are asked to predict where Max will look for his chocolate based on his beliefs about its location (Perner & Lang, 1999; Wimmer & Perner, 1983).

Another well-known explicit false belief task is the content false belief task (Hogrefe et al., 1986; Perner et al., 1987), also named the *"Smarties task"* and the *deceptive-container task*. In this task, children are shown a container, such as a Smarties box, and asked, "What is in here?" The children typically answer "Smarties". The box is then opened, revealing that there is actually a pig figurine in the box. The pig figurine is then put back in the box, and the box is closed. Participants are then asked two questions: a control question ("Can you remember 'what is inside here?") and a false belief question: "Lukas (another child or a toy figurine) has never seen 'what is in this box. What would he think is in the Smarties box?". Regarding the percentage of correct responses to the belief question across age groups (see **Figure 3**), the ability to attribute a false belief was nearly absent among 3-year-olds. Approximately 50% of 4-year-olds and 80% of 5-year-olds provided correct answers. To answer correctly, the child must meta-represent how the world previously appeared to them (albeit incorrectly) or how it would appear to another individual (Rakoczy, 2022). By age 4, children succeed in such tasks by attributing to both their earlier self and another agent the false belief that the Smarties box contained a pig figurine (Gopnik & Astington, 1988; Rakoczy, 2022).

Standard explicit elicited-response false belief tasks are by far the most frequently used and the most richly interpreted measures of Theory of Mind (Apperly et al., 2011). These tasks involve individuals holding beliefs about a target object that differ from reality, effectively distinguishing children's reliance on mental state reasoning from their use of reality-based behavioral predictions, which necessitates comprehension of representational change (Gopnik & Astington, 1988; Perner, 1991) and the ability to make the appearance-reality distinction (Flavell, 2000; Flavell et al., 1983). As such, they have been regarded as the gold standard or "litmus test" for assessing Theory of Mind abilities. In all of these tasks, children were introduced to a protagonist who lacked crucial information, resulting in a false belief about the situation, unlike the child. The child was then explicitly asked either to describe the protagonist's belief (e.g., "What does the protagonist think?") or to predict the protagonist's actions (e.g., "Where will the protagonist look for the target object?"). Across different variations of these explicit false belief tasks, findings from this line of research consistently revealed that children older than 4 years are able to complete explicit false belief tasks as they start to acquire a proper notion of mental representation. In contrast, children younger than 3½ years typically fail to do so (Wellman et al., 2001). Specifically, in the location false belief task, children under 3 years old say that the protagonist thinks the object is where it really is and accordingly say that the protagonist will search in the target object's actual location (i.e., blue box in Max's chocolate task). Similarly, in the content false belief task, children erroneously assert that an uninformed agent will believe that the contents of a Smarties box are what they actually are (i.e., a pig figurine) rather than Smarties. The blue bars in **Figure 3** illustrate the percentage of children who provided correct answers on the false belief tasks. The developmental progression indicated by these figures suggests that children begin to understand an explicit false belief as a representational state of mind around the age of 4.

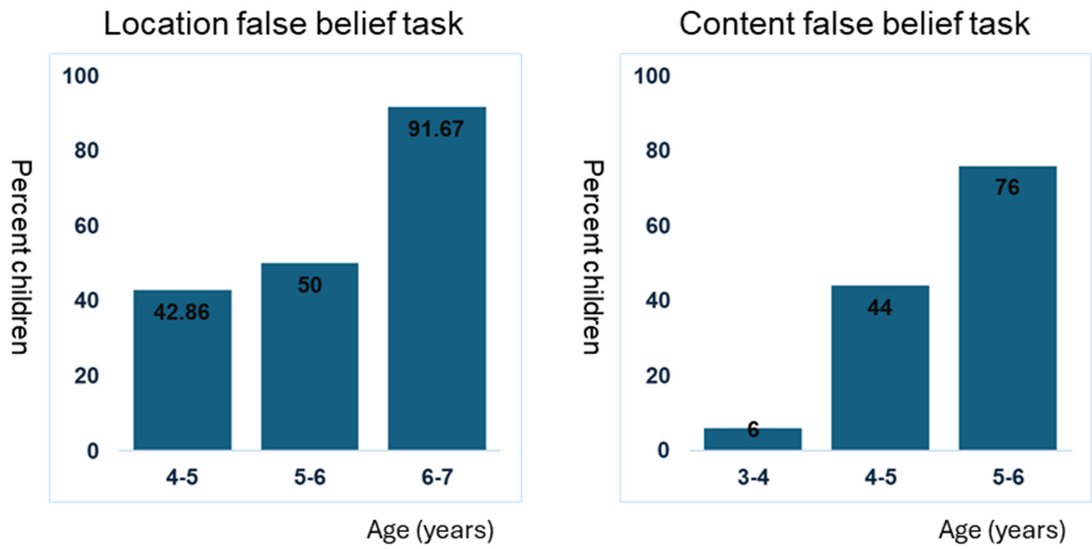


Figure 3. Correct answers to belief questions. Left figure: Data sourced from Wimmer and Perner (1983). The data for 4- to 6-year-olds are from Experiment 2, while those for 6- to 7-year-olds are from Experiment 1. Right figure: Data sourced from Hogrefe et al. (1986). The data for 3- to 5-year-olds are from Experiment 1.

The initial findings of a marked age-related trend in false belief understanding prompted a substantial body of research investigating the effects of task modifications, including verbal demands, the salience of reality, the wording of tests, and question formats. For example, Mitchell and Lacohée (1991) conducted a study wherein children mailed a picture of the tube to a postbox upon the first presentation of a Smarties tube. When subsequently queried, most 3-year-olds accurately stated their initial belief and the tube's actual contents. Additionally, Kikuno et al. (2007) found that altering the target object's identity during its transfer (e.g., change from a **hat** to an **apple**) and querying children about the original object in the belief question (i.e., "Where will the agent look for his **hat**?") significantly increased the percentage of correct responses among 3-year-olds. Furthermore, temporally marking the test question (e.g., "Where will the protagonist **look first** for the target object?") improved children's performance compared to standard questions, resulting in either above-chance or chance-level outcomes (e.g., Surian & Leslie, 1999; Yazdi et al., 2006). More recently, Ghrear et al. (2021) revealed that minimizing the curse of knowledge by removing specific outcome information allowed 3-year-olds to perform on par with older preschoolers in tasks assessing their reasoning about others' false beliefs. Despite these facilitative effects observed in individual studies through various task modifications (e.g., Bartsch, 1996; Ghrear et al., 2021; Kikuno et al., 2007), reducing the processing demands sometimes enabled 3-year-olds to succeed, though they often only performed at chance level. Nevertheless, a meta-analysis conducted by Wellman et al. (2001) found that false belief performance exhibited a consistent developmental trend between the ages of 2½ and 5 years. In Wellman et al.'s (2001) meta-analysis, while some studies noted improved performance levels in 3-year-olds, these children did not generally outperform chance on false belief tasks. By around 3½ years of age, approximately half of the children succeeded in such tasks, with a marked performance improvement thereafter. By the age of 4 years, children consistently outperformed the chance threshold of 50% correct responses. Thus, around age 4 represents a sensitive period for Theory of Mind, where children start correctly reasoning about others' false beliefs across different task settings and belief contents (Flavell et al., 1983, 1990; Hogrefe et al., 1986; Wellman et al., 2001; Wimmer & Perner, 1983). Detailed discussions of various theoretical perspectives on the emergence of Theory of Mind (e.g., Bartsch & Wellman, 1995; Perner, 1991) are elaborated in Section 1.1.2.

Low-demands explicit false belief task

While the previously described task modifications improved the performance of 3-year-olds from below-chance to chance level, these explicit false belief tasks involve at least three cognitive processes (Setoh et al., 2016). The first is false belief representation: As the story unfolds, children must construct and maintain a representation of the protagonist's false belief. The second is response generation: children must interpret the question, retain it, and generate an appropriate response. The third is inhibitory control: Children must inhibit the incorrect prepotent response to access their representation of the protagonist's false belief and provide the correct answer. According to Setoh et al.'s (2016) perspective, young children may fail these tasks for two reasons. One possibility is that the degree of inhibitory control required is beyond the capability of many children, leading to chance performance as some succeed while others do not. Alternatively, young children may generally have sufficient inhibitory control but struggle with the concurrent processing demands, overwhelming their limited information-processing resources and resulting in confused or random responses from the children. Recently, Setoh et al. (2016) conducted four experiments demonstrating that 30- to 33-month-old toddlers can succeed in an explicit false belief task when overall processing demands—including response-generation and inhibitory-control demands—are decreased. This finding strongly supports the view that early difficulties with standard explicit false belief tasks arise from their high processing demands, as even toddlers can succeed in such tasks when these demands are reduced.

In Experiment 1 (see **Figure 4**), participants were introduced to a scenario involving a protagonist named Emma, who initially found an apple in one of two containers on a table: a bowl obscured by a towel to the left and a lidded box to the right. Emma then moved the apple to the other container. Following this, she left to play outside with her ball. During her absence, her brother Ethan found the apple and took it away. Emma then returned to look for her apple. In the test trial, children saw pictures of the bowl and box and were asked, “Where will Emma look for her apple?” For success in the task, children were required to identify the container where Emma mistakenly believed her apple was located. Two practice trials were included between the story trials to reduce the response-generation demands during the test trial. During each practice trial, children were shown two pictures and asked a “where” question, similar to the test trials. In one practice trial, toddlers saw an apple and a banana and were asked, “Where is Emma’s apple?” In the other practice trial, they saw a ball and a frisbee and were asked, “Where is Emma’s ball?”. The test object (i.e., apple) was removed from the scene to reduce demands for inhibitory control, obviating the requirement for children to overcome “reality bias” or the “curse of knowledge,” which has been assumed to hinder younger children’s successful false belief representation (Kaltefleiter, 2022; Robinson & Mitchell, 1995). Under these experimental configurations, 30- to 33-month-old toddlers performed above chance in the test trial; 25 of 32 (78%) toddlers pointed to the false belief container (Setoh et al., 2016, Exp. 1).

To determine the essential features of practice trials for success in a false belief task, Setoh et al. conducted Experiment 2 and Experiment 3. Experiment 2 modified the practice trials to include only one picture instead of two. The results showed that 30-month-old toddlers failed under these conditions, underscoring the importance of aligning practice trials with test trials to support their comprehension. Experiment 3 further investigated whether both “where” questions were necessary to significantly lower the response-generation demands of the test trial. The findings revealed that performance among 33-month-old toddlers dropped to chance levels when only one “where” question was presented, indicating that exposure to two such questions—one in each practice trial—was crucial for their success in the false belief task. These results suggest that toddlers succeeded in a traditional elicited-response false belief task only when the type and amount of practice provided were sufficient to reduce the response-generation demands of the test trial to a manageable level.

Two modifications were implemented in Experiment 4 to investigate the impact of inhibitory control demands on children’s performance. First, rather than removing the apple, Emma’s brother relocated it from where he initially found it to the other container before leaving. This modification created a high-inhibition task, as toddlers knew the apple’s actual location. During the response phase, the actual location of the target object is assumed to trigger a “pull of the real,” which increases the task’s inhibitory demands and complicates the successful completion of the task (Kaltefleiter, 2022). Second, researchers counterbalanced whether Emma found and hid the apple in the same container (one-transfer condition) or in different containers (two-transfer condition). In the test trial, only 7 out of 32 children (22%) correctly pointed to the container corresponding to Emma’s false belief. The negative results of Experiment 4 confirmed that reducing the response-generation demands without addressing the inhibitory-control demands resulted in below-chance performance. This result was likely due to children’s limited inhibitory control, which prevented them from suppressing the strong, incorrect prepotent response elicited by the standard question.

Setoh et al.'s (2016) study suggests that young children's failures in explicit false belief tasks are mainly due to limited information processing abilities, especially when they need to inhibit a prepotent response and generate a verbal answer to an unfamiliar "where" question (Sodian et al., 2024). As predicted, toddlers performed above chance on a traditional false belief task after reductions in inhibitory control demands and the addition of two response-generation practice trials. This finding was replicated by an independent laboratory (Grosso et al., 2019), which reported a correct response rate of 74% among *N* = 58 toddlers aged 33 months. Additionally, the original authors of the low-demands false belief task found that 2.5-year-old toddlers could succeed in elicited-response false belief tasks involving false beliefs about object location or identity when the processing demands were sufficiently reduced (Scott et al., 2020). Additionally, a recent longitudinal study found a significant correlation between false belief understanding assessed using the low-demands task by Setoh et al. (2016) at 33 months and performance on a standard false belief assessment at 52 months (Sodian et al., 2024). This finding further provides evidence against a low-level associationist interpretation of performance on the low-demands false belief task. All these findings suggest that children are capable of attributing genuine false beliefs prior to 4 years of age, and explicit false belief understanding may develop continuously from toddlerhood to childhood (Scott et al., 2022).

Figure 4. Apparatus, scripts, and pictures were used in the low-demands false belief task (Setoh et al., 2016).

1.2.1.2. Behavioral evidence of implicit false belief understanding

In traditional explicit elicited-response false belief tasks, children are directly questioned about what an agent believes or how he will act (Baron-Cohen et al., 1985; Gopnik & Astington, 1988; Perner et al., 1987; Wimmer & Perner, 1983). These tasks require an explicit judgment that represents the child’s understanding of another’s beliefs. However, novel implicit spontaneous-response false belief tasks revealed that even infants in their first year can exhibit varying expectations or interpretations of an agent’s actions based on his beliefs (e.g., Baillargeon et al., 2010; Martin & Santos, 2016). In these implicit false belief tasks, children’s understanding of an agent’s false belief is inferred from their spontaneous behaviors as they observe a scene unfold (Baillargeon et al., 2010, 2015). Children may possess an implicit understanding of false beliefs that can still guide their actions while not subject to conscious deliberation (Dienes & Perner, 1999). This understanding influences their choices, leading to appropriate decisions despite their inability to verbalize these beliefs (Vierkant, 2012).

In 1994, Clements and Perner employed an anticipatory looking task within the framework of a standard unexpected-transfer false belief paradigm, where the experimenter mused to himself, “I wonder where she is going to look?” instead of directly asking the children. This prompt was designed to elicit anticipatory gaze behavior based on the agent’s belief. Most 35-month-old children correctly anticipated the agent’s search location based on her belief, as evidenced by their initial gaze before responding to test questions (Clements & Perner, 1994). According to Clements and Perner (1994), the anticipatory looking responses provided evidence of an implicit understanding of belief. Garnham and Ruffman (2001) presented 2- to 4-year-old children with a false belief story similar to that in Clements and Perner’s (1994) study, but with three containers instead of two. The protagonist initially placed the object in the left-hand box, but another character moved it to the right-hand box in his absence. If children's anticipatory gaze were guided by belief sensitivity, they would look at the left-hand box. In contrast, an association bias or ignorance sensitivity would lead them to look at either the left-hand or middle box. The results supported the former, suggesting that the implicit understanding of belief influences children's anticipatory looking.'

Clements and Perner (1994) pioneered the measurement of children's understanding of implicit false beliefs through an analogous spontaneous-response paradigm with video camera recording. Building on these seminal findings regarding two-year-olds' implicit understanding of false belief, numerous subsequent studies employing spontaneous-response paradigms through eye-tracking techniques have significantly enriched our understanding of the scope and depth of infants' and young toddlers' implicit knowledge of false beliefs (e.g., Buttelmann et al., 2009; Onishi & Baillargeon, 2005; Southgate et al., 2007). The spontaneous-response paradigms predominantly included anticipatory-looking, violation-of-expectation, and prompted action paradigms, which are detailed below.

Anticipatory looking paradigm

Anticipatory looking paradigms make use of the innate human propensity to predict actions and their consequences, a tendency that emerges by the end of the first year of life (Falck-Ytter et al., 2006; Flanagan & Johansson, 2003). The paradigm involves presenting actions based on specific mental states—such as intentions or true/false beliefs—and observing participants' anticipatory gaze to determine whether they consider the agent's mental state when predicting their action (Clements & Perner, 1994; Kaltefleiter et al., 2022; Schneider et al., 2012; Schuwerk et al., 2018; Senju et al., 2011; Southgate et al., 2007; Surian & Geraci, 2012; Thoermer et al., 2012; L. U. Wang & Leslie, 2016). Building on prior anticipatory looking results with 3-year-olds (Clements & Perner, 1994; Garnham & Ruffman, 2001), Southgate et al. (2007) refined the anticipatory looking methodology using advanced eye-tracking technology by eliminating verbal interaction and removing the object from the scene to test 25-month-old infants in a spontaneous-response false belief task. In Southgate et al.'s (2007) anticipatory looking paradigm, the familiarization trials were designed to convey that agent's goal was to retrieve the hidden object, as well as the general timing and structure of events, to set up participants' expectations in the test trials. The target object was placed in the left-hand box, observed by the infant, and retrieved by the agent from the same box. The subsequent trial followed a similar procedure but with the object placed in the right-hand box, ending as the agent made contact with the box. Successful anticipation of which door the agent would open during the second trial was required for infants to proceed to subsequent test trials. The test trials, categorized as False-Belief 1 (FB1) and False-Belief 2 (FB2), aimed to determine whether infants make anticipatory saccades based on an attribution of false belief. In the FB1 condition, an agent witnessed the target object being moved from box A to box B but was absent when it was away, resulting in a false belief that the object was in box B. In the FB2 condition, the agent missed both the transfer from box A to box B and the subsequent removal, leading to a false belief that the object remained in box A. The FB1 and FB2 conditions serve as mutual controls to confirm that children genuinely track beliefs to complete the task (Baillargeon et al., 2018; Kulke, Wübker, et al., 2019; Southgate et al., 2007). Southgate et al. (2007) observed that most 25-month-old children initially directed their gaze to the correct belief-based location and spent more time observing it compared to other locations.

Replications of findings in anticipatory looking tasks yielded mixed results, including partial replications (Grosse Wiesmann et al., 2017a; Kaltefleiter et al., 2022; Schneider et al., 2012; Surian & Geraci, 2012) and failures to replicate (Kampis et al., 2021; Kulke, Reiß, et al., 2018; Kulke, von Duhn, et al., 2018; Kulke, Wübker, et al., 2019; Schuwerk et al., 2018). The lack of a clear pattern from these non-replications may be due to the wide variety of methodological parameters used across studies, including differences in participants' ages, stimuli, and analysis criteria. Given the mixed and complex evidence, it remains unclear whether infants and toddlers engage in implicit false belief understanding (e.g., Dörrenberg et al., 2018) and necessitates a systematic, large-scale, multi-lab study designed to rigorously test the robustness, reliability, and replicability of implicit false belief understanding measures (Steffan et al., 2024). The *ManyBabies 2* project (Schuwerk et al., 2024; Zimmer et al., 2024) is a significant initiative aimed at systematically assessing the reliability of findings from the spontaneous-response anticipatory looking paradigm.

Violation-of-expectation paradigm

Violation-of-expectation looking-time paradigm (Kovács et al., 2010; Onishi & Baillargeon, 2005; Scott et al., 2012; Scott & Baillargeon, 2009; Surian et al., 2007; Yott & Poulin-Dubois, 2012) is derived from habituation paradigms. Infants typically observe a scene wherein an agent forms a mental state, such as a belief, which may be true or false. The infants are then shown the agent acting either in accordance with this belief (expected behavior) or contrary to it (unexpected behavior). The key measure here is the duration of the infants' gaze: longer looking times at unexpected outcomes indicate surprise or increased interest, suggesting that the infants had formed implicit expectations about the agent's actions. This prolonged attention to the unexpected outcome is supposed to indicate the infant's emerging ability to consider others' false beliefs when predicting their actions. For example, in Onishi and Baillargeon's (2005) study, infants were initially exposed to three familiarization trials, where they observed an agent interacting with a toy and two boxes, one green and one yellow. The agent placed the toy in the green box and later grasped it in subsequent trials. The subsequent belief conditions aimed at establishing whether the agent held a true (TB) or false belief (FB) about the toy's location. TB conditions: The agent observed the toy either remaining in box A or witnessed its transfer to box B. FB conditions: Unbeknownst to the agent, the toy was moved to box B, or the agent saw the toy being moved to box B but failed to observe its return to box A. In the test phase, the actor opened the doors and reached into the boxes A or B. Fifteen-month-old infants looked significantly longer when the agent searched in a location inconsistent with her belief about the toy's location (Onishi & Baillargeon, 2005). These results indicate that 15-month-olds may already have a representational Theory of Mind, recognizing that others act based on beliefs, which are representations that may or may not reflect reality (Onishi & Baillargeon, 2005).

Replication of findings in violation-of-expectation tasks was mixed, encompassing successful replications (e.g., Song et al., 2008; Surian & Geraci, 2012; Yott & Poulin-Dubois, 2012), partial replications (e.g., Dörrenberg et al., 2018), and non-replications (e.g., Poulin-Dubois et al., 2013, 2020; Poulin-Dubois & Yott, 2018; Powell et al., 2018; Yott & Poulin-Dubois, 2016). Negative findings with children suggest that seemingly minor procedural differences can significantly impact results. For instance, a closely aligned replication by Surian et al. (2007) provided evidence that even 13-month-olds can attribute to an agent a false belief about the location of an object. Another replication by Yott and Poulin-Dubois (2012) also corroborated these findings, affirming the robustness of the original experimental design. In contrast, Poulin-Dubois et al. (2013) introduced modifications by making the boxes transparent and adding a fourth familiarization trial in which the agent ignored the toy and donned a blindfold that she also wore during the test phase. These alterations created an ambiguous scenario that likely hindered children's ability to form clear expectations about the blindfolded agent's subsequent actions, which may have contributed to the negative results observed.

Prompted action paradigm

Prompted action task is another type of implicit false belief task that takes advantage of children's spontaneous or indirectly prompted behavior in response to the actions of an agent with false beliefs, as well as their propensity to assist the agent in achieving these goals (Buttelmann et al., 2009; Rakoczy, 2022). The agent forms a mental state (e.g., a true or false belief about a box's contents), and experimenters assess whether participants spontaneously consider the agent's belief in their interactions (e.g., by helping or interpreting the agent's communicative actions) (Buttelmann et al., 2009, 2014; Knudsen & Liszkowski, 2012; Southgate et al., 2010). For instance, in Buttelmann et al.'s (2009) study, experimenter 1 (E1) showed 18-month-olds how to lock and unlock box A and B; the child had successfully opened each box twice in turn without E1's help. Next, experimenter 2 (E2) entered the room and excitedly showed the caterpillar toy to the child, , hid the toy in box A, and then left. In the false belief condition, while E1 was absent, E2 moved the toy to box B and locked both boxes. Upon E1's return, he unsuccessfully tried to open box A. When prompted to help, most infants approached box B, indicating they understood that E1 falsely believed the toy was still in box A. In another similar task by Southgate et al. (2010), two toys were initially placed separately in boxes A and B, and their locations were subsequently switched under false belief conditions. Despite this modification, 17-month-old infants were still able to track the agent's epistemic state and use this information to infer the agent's intended referent.

The replication of findings in prompted action tasks also yielded mixed results, including successful (Király et al., 2018; Southgate et al., 2010) and failed replications (Dörrenberg et al., 2018; Poulin-Dubois et al., 2023; Wenzel et al., 2020). For the negative findings with children, some minor procedural differences can lead to negative results, including differences in materials and procedure, interference effect, and non-standard criteria for ending each test trial. For example, Crivello and Poulin-Dubois (2018) introduced “slight methodological changes” in their replication of Buttelmann et al.’s (2009) study, which failed to replicate the original findings. Differences in setup and procedure may have contributed to this failure. To reduce the drop-out rate, the experimenters seated participants closer to the test boxes, allowing for ambiguous choices and reducing the time available to process observations before responding, potentially leading to chance-level performance between conditions.

Replication crisis of implicit false belief understanding

Results from nontraditional spontaneous-response false belief tasks steadily accumulated over the past 20 years. Some studies with infants and children successfully replicated the original findings, while others only partially replicated them or failed to replicate them entirely, thereby calling into question the reliability of these tasks (Kulke, Johannsen, et al., 2019; Kulke & Rakoczy, 2018; Rakoczy, 2022). The discrepancy between positive original findings and negative replications in implicit false belief studies may result from differences in experimental procedures, task design, sample sizes, and measurement methods (Baillargeon et al., 2018).

Since the publication of the original implicit false belief tasks, evidence from multiple studies involving infants and children attempting to replicate the original findings was complex and mixed (see **Table 2**). Given that the most substantial evidence for implicit false belief understanding comes from studies utilizing anticipatory looking paradigms, it is essential to address the replication crisis by developing a more robust anticipatory looking paradigm. The paradigm proposed by Grosse Wiesmann et al. (2017a) represents a promising step toward addressing these challenges. Grosse Wiesmann et al. (2017a) created a multi-trial anticipatory-looking paradigm involving FB1 and FB2 conditions to reduce high exclusion rates resulting from incorrect anticipatory responses. The experimental condition featured two animated agents engaging in a chase, with one agent hiding from the other in a Y-shaped tunnel (Surian & Geraci, 2012). The design of the FB1 and FB2 conditions was similar to that of Southgate et al. (2007). Using this refined paradigm, Grosse Wiesmann et al. (2017a) demonstrated above-chance performance in both FB1 and FB2 conditions among 3- and 4-year-old children.

A subsequent study following a similar approach to Grosse Wiesmann et al. (2017a) found that all age groups (children aged 27, 36, and 52 months) performed significantly above chance in the FB1 condition but significantly below chance in the FB2 condition, along with prolonged looking times at the final hiding location (Kaltefleiter et al., 2022). In the FB2 condition, children may fail to consider that the agent did not witness the target's transfer, instead focusing on the last location they observed the target themselves, disregarding the agent's lack of this information (Kaltefleiter et al., 2022). Moreover, the FB2 condition may place greater demands on working memory, attention, and inhibitory control as children must track the target's actions, represent the agent's belief, and sustain this belief over time against their own conflicting perspective (Baillargeon et al., 2018; Grosse Wiesmann et al., 2018; Kaltefleiter et al., 2022). The same pattern of above-chance FB1 and below-chance FB2 performance has also been observed in other anticipatory looking studies (Dörrenberg et al., 2018; Grosse Wiesmann et al., 2018; Kampis et al., 2021; Kulke et al., 2018, Study2b). These findings suggest that success in the FB1 condition may reflect competencies in mental state reasoning similar to those assessed by other implicit false belief tasks, whereas the FB2 condition, due to its higher cognitive demands, may not reliably measure implicit false belief understanding (Baillargeon et al., 2018; Kaltefleiter et al., 2022). If the FB1 condition in the anticipatory looking paradigm reliably measures implicit false belief understanding, a critical question arises: Is there a corresponding neural correlate underlying the behavioral competence exhibited in the FB1 condition?

Table 2. Overview of measures of implicit false belief understanding and the original and replication findings, adapted from (Rakoczy, 2022).

Measure	Original findings	Reproducibility of original findings
Anticipatory looking task	Infants accurately anticipated an agent's actions based on the agent's belief (Southgate et al., 2007)	Mixed, including partial (Grosse Wiesmann et al., 2017a; Kaltefleiter et al., 2022; Schneider et al., 2012; Surian & Geraci, 2012) and non-replications (Kampis et al., 2021; Kulke, Reiß, et al., 2018; Kulke, von Duhn, et al., 2018; Kulke, Wübker, et al., 2019; Schuwerk et al., 2018)
Violation-of-expectation task	Infants consistently looked longer when the actor's search location contradicted the agent's belief about the toy's location (Onishi & Baillargeon, 2005)	Mixed, including successful(Song et al., 2008; Surian & Geraci, 2012; Yott & Poulin-Dubois, 2012), partial (Dörrenberg et al., 2018) and non-replications (Poulin-Dubois et al., 2020; Poulin-Dubois & Yott, 2018; Powell et al., 2018; Yott & Poulin-Dubois, 2016)
Prompted action task	Infants successfully considered the adult's belief when attempting to infer his goal (Buttelmann et al., 2009)	Mixed, including successful (Király et al., 2018; Southgate et al., 2010) and failed replications (Dörrenberg et al., 2018; Poulin-Dubois et al., 2023; Wenzel et al., 2020)

1.2.2. Theoretical perspectives on the relationships between implicit and explicit false belief understanding

Early research on Theory of Mind indicated that children under four typically fail explicit false belief tasks. However, implicit and low-demands explicit false belief tasks reveal that even infants and toddlers can attribute false beliefs to others. This discrepancy between implicit and explicit task performance necessitates further investigation into why children under four struggle with explicit tasks while infants succeed in implicit ones, and the cognitive processes underlying these differences. The relationship between implicit and explicit false belief understanding and the extent to which implicit false belief understanding reflects a complete and flexible representation of others' mental states is a matter of debate (Scott et al., 2022; Sodian, 2016). Over the past 40 years, our understanding of the development of false belief underwent several substantial shifts, prompted by the previously unsuspected false belief competencies in novel implicit false belief tasks. Several theoretical perspectives were proposed to synthesize these findings and construct frameworks concerning the potential relationship between implicit and explicit false belief understanding (see reviews by Carruthers, 2013; Scott et al., 2022; Sodian et al., 2020).

Earlier accounts: Conceptual change accounts

Earlier accounts, such as conceptual change accounts (see Section 1.1.2 for a detailed overview of the theoretical perspectives on the emergence of Theory of Mind), support the perspective that children under the age of 4 are incapable of representing false beliefs (Baron-Cohen et al., 1985; Gopnik & Astington, 1988; Perner, 1991; Perner et al., 1987; Wellman, 1990; Wimmer & Perner, 1983). In the traditional location false belief task (Wimmer & Perner, 1983), children attribute to the agent a subjective misrepresentation: a false belief that deviates from reality and is incompatible with the child's own perspective. Theory of Mind representation requires the establishment of a situational context and narrative to convey the agents' understanding of the world, as exemplified by the classic unexpected transfer task, such as *Max's chocolate story* (Wimmer & Perner, 1983; see Section 1.2.1.1). Successfully pointing to the original location of the object, rather than the new one is considered a key behavior marker for the emergence of explicit false belief understanding.

The developmental trajectory of explicit false belief understanding led proponents of conceptual change theories to assert that a fundamental transformation in psychological reasoning occurs around age 4 (Baron-Cohen et al., 1985; Liu et al., 2008; Wellman et al., 2001). Within this age range, children begin to develop meta-representations necessary for reasoning about others' false beliefs by representing mental states while simultaneously maintaining an accurate understanding of reality (Perner, 1991). Perner explains that meta-representation involves understanding both the content of a representation (e.g., the chocolate) and its mental representation (e.g., Max's mental image of the chocolate), as well as the coreferential representational relationship between them (i.e., Max's mental image represents the actual chocolate) (Grosse Wiesmann, 2018; Perner, 1991). He contends that children under the age of 4 may lack the ability for meta-representation and thus fail to recognize that others' mental representations may differ from their own. For instance, in a location false belief task, the child needs to meta-represent how the agent represents the location of the target object. In a content false belief task, the child has to meta-represent how the world (wrongly) appeared to them previously or would appear to another person (Rakoczy, 2022). Conceptual change, in addition to representing a transition from a situation-based to a representation-based understanding of behavior (Perner, 1991), also involves a shift from primitive, desire-based psychology to more sophisticated belief-desire psychology (Bartsch & Wellman, 1995; Wellman, 1990), marking a critical Theory of Mind developmental milestone. Mastery of false belief understanding is therefore recognized as a significant achievement in early childhood, marking the advent of advanced psychological reasoning (Carlson

& Moses, 2001; de Villiers & de Villiers, 2003; Gopnik & Wellman, 1994; Perner, 1991).

However, conceptual change theories have been challenged by new evidence suggesting that even infants can demonstrate an understanding of false beliefs when tested using implicit spontaneous-response false belief tasks, including anticipatory looking tasks (Senju et al., 2011; Southgate et al., 2007; Surian & Geraci, 2012; Thoermer et al., 2012), violation-of-expectation tasks (Knudsen & Liszkowski, 2012; Moll et al., 2017; Onishi & Baillargeon, 2005; Scott et al., 2012; Southgate et al., 2010), and prompted action tasks (Buttelmann et al., 2009; Southgate et al., 2010). Findings from these implicit false belief tasks suggest that some ability to reason about false beliefs may already be present early in life. However, the extent to which these implicit false belief tasks reflect an understanding of an agent's mental representation remains unclear. Infants may succeed in these tasks by adopting the agent's representation without independently maintaining their own accurate representation of reality (Kulke, Johannsen, et al., 2019; Poulin-Dubois et al., 2023; Rakoczy, 2022). Thus, the nature of infants' understanding of false beliefs and the relationship between implicit and explicit false belief comprehension remain subjects of ongoing debate.

Conceptual-continuity accounts

Traditional explicit tasks indicate that false belief understanding emerges around age 4, marking a significant milestone in Theory of Mind. However, findings from implicit spontaneous-response tasks accumulated over the past two decades suggest that infants (under two years) and toddlers (ages 2–3) already demonstrate false belief understanding, indicating its early integration into Theory of Mind. According to conceptual continuity accounts (also referred to as s*ubstantial-continuity accounts* or *one-system accounts*), implicit and explicit false belief tasks tap the same genuine false belief understanding, but explicit false belief tasks are subject to greater processing difficulties, such as inhibitory control demands and response generation demands (Baillargeon et al., 2010; Scott, 2017; Scott et al., 2022). Inhibitory control requires children to suppress an initial, incorrect response to accurately reflect the protagonist’s false belief. Response generation involves children interpreting, retaining, and responding to a standard question that predicts the protagonist’s behavior.

Conceptual continuity theory predicts cross-sectional and longitudinal interrelations among measures of mental state attribution that are conceptually related (Sodian et al., 2020). These relationships are expected to be independent of general cognitive functioning or language ability. In Setoh et al.’s (2016) study, children aged 30 to 33 months were shown an explicit false belief task with reduced processing demands. Toddlers performed above chance in the test trial; 25 of 32 (78%) children pointed to the container that Emma mistook to hold her apple (Setoh et al., 2016, Exp. 1). Above-chance performance in Setoh et al. (2016) task was replicated in another independent laboratory (Grosso et al., 2019) and further extended to encompass false belief understanding about identity in toddlers aged 2.5 years (Scott et al., 2020).

To date, a comprehensive longitudinal study, “Theory of Mind in Infancy and Early Childhood” (TOMII/TOMEC), found strong longitudinal relationships, independent of verbal IQ, between an implicit false belief task using anticipatory looking at 18 months and explicit false belief tasks conducted annually between ages 4 and 6, as well as a belief-based intention task (Kloo et al., 2020; Sodian et al., 2016, 2020; Thoermer et al., 2012). Moreover, another independent study also found a longitudinal significant correlation between false belief understanding assessed with the low-demands task by Setoh et al. (2016) at 33 months and false belief understanding in a standard content false belief assessment at 52 months (Sodian et al., 2024).

Furthermore, conceptual continuity theory predicts relationships between theoretically relevant behaviors in infancy, such as goal-directed action (Sodian, 2011; Sodian & Kristen-Antonow, 2015) and later explicit Theory of Mind, independent of general cognitive functioning. Prior research identified links between early goal-directed attention and subsequent Theory of Mind abilities (Aschersleben et al., 2008; Peterson et al., 2005; Wellman, 2002; Wellman et al., 2004, 2008; Wellman & Liu, 2004; Yamaguchi et al., 2009). For example, a longitudinal correlation showed that 6-month-olds' attention to goal-directed actions predicted false belief task performance at age 4, independent of language skills (Aschersleben et al., 2008). Similarly, habituation to intentional actions at 14 months predicted preschool mentalistic reasoning (Wellman et al., 2004), and social attention at 11 months predicted later false belief understanding (Wellman et al., 2008). These findings indicate that false belief understanding is grounded in early goal-directed attention (Moll & Meltzoff, 2012; Sodian et al., 2020).

These findings mentioned above support the view of conceptual continuity from preverbal psychological reasoning in infancy to explicit verbal reasoning about mental states at 4–5 years (Sodian et al., 2020). False belief understanding development seems to involve the enrichment and increasing stability of core conceptual understanding, rather than fundamental shifts (Sodian, 2004; Sodian et al., 2024). These findings also lend support to dual-system theory, which, along with developmental enrichment accounts, will be discussed in subsequent sections.

Two-system accounts

Two-system accounts grant infants a minimal capability for reasoning about mental states and assume that implicit and explicit false belief understanding develop independently (Apperly & Butterfill, 2009; Butterfill & Apperly, 2013; De Bruin & Newen, 2012; Fizke et al., 2017; C. D. Frith & Frith, 2008; Grosse Wiesmann et al., 2020; Low et al., 2016; Low & Watts, 2013; Sodian, 2011; Wellman, 2014). They propose that success at implicit false belief tasks is distinct from, and considerably more primitive than, the late-developing system that emerges around age 4 and enables success at explicit false belief tasks. Aligned with dual-process theories in other cognitive domains (e.g., Evans & Stanovich, 2013), an early-developing, less cognitively demanding, automatic, and possibly implicit system coexists with a later-developing, more demanding, yet flexible system. The early system allows infants to predict false belief scenarios correctly, while the later system facilitates flexible verbal attribution of false beliefs across various situations.

The early-developing system for implicit false belief understanding does not represent false beliefs directly; instead, it tracks belief-like registrations. When an agent encounters an object, its location and properties are registered. By tracking this registration—even if it becomes outdated during the agent's absence—the early system can predict the agent's actions, such as searching for a toy where it was last registered (Low et al., 2016; Scott et al., 2022). Butterfill and Apperly (2013) argued that infants do not track propositional attitudes (e.g., beliefs) but track "simpler, relational mental states" that correspond with propositional attitudes. Implicit false belief understanding develops early, operates unconsciously, and functions in spontaneous and indirect tasks. It involves a basic Theory of Mind, which represents mental states beyond perception-goal reasoning but does not yet include belief-desire psychology. In contrast, explicit false belief understanding—representing a fully developed meta-representational Theory of Mind—emerges later, relying on language and executive function, and operates explicitly and consciously.

In the TOMII/TOMEC study, implicit and explicit false belief tasks showed different relationships with executive functions (Sodian et al., 2020). Explicit false belief understanding was found to be associated with executive functions, while implicit false belief understanding showed no correlation with executive function tasks administered between ages 2 and 5 (Sodian et al., 2020). Moreover, the correlation between explicit false belief understanding and executive functioning was independent of implicit false belief understanding (Kloo et al., 2020). Grosse Wiesmann et al. (2017a) also found that explicit false belief performance correlated with executive functions and syntactic abilities, whereas implicit false belief did not. The clear differentiation between implicit and explicit false belief processing concerning executive functions challenges the idea of a single neurocognitive mechanism instead of supporting a moderate two-system account (Sodian et al., 2020).

The two-system account is also supported by evidence from the neural level. A study by Grosse Wiesmann et al. (2020) using the functional magnetic resonance imaging (fMRI) technique found that explicit false belief understanding was associated with cortical surface area and thickness in several right-hemispheric regions, including the precuneus, posterior middle temporal gyrus (MTG), and TPJ. In contrast, implicit false belief understanding was linked to a distinct neural network, particularly the right supramarginal gyrus (SMG). These findings suggest the possibility of two systems for reasoning about mental states: a more mature, explicit false belief understanding emerging around age 4, and an earlier-developing implicit false belief understanding. Grosse Wiesmann et al. (2020) and other studies (e.g., Bardi et al., 2016; Boccadoro et al., 2019; Hyde et al., 2018) provide preliminary evidence that right-hemispheric regions are involved in Theory of Mind, with distinct areas supporting implicit and explicit false belief understanding. Further research, particularly in younger children, is needed to clarify the developmental trajectories of these neural systems.

Developmental enrichment accounts

In addition to conceptual continuity accounts and dual-system theories, a moderate perspective posits that developmental enrichment occurs between infants' implicit false belief understanding and explicit false belief understanding at 4 years of age (e.g., Symeonidou et al., 2024). Developmental enrichment theory is a heterogeneous category of theory, suggesting that the Theory of Mind originates in infancy and involves developmental change. While infants do not possess a fully developed Theory of Mind, even in its rudimentary or preconceptual forms, they do possess foundational abilities—such as statistical learning (Ruffman et al., 2012) and shared intentionality in joint attention (Tomasello, 2018) — that establish the basis for the later development of a Theory of Mind.

Ruffman and colleagues (Ruffman, 2014; Ruffman et al., 2012) propose that statistical learning serves as a foundation for Theory of Mind in children. Infants possess an innate ability for statistical learning, and their interest in faces and animate motion, combined with maternal input, enables them to learn about mental states (Ruffman et al., 2012). This early ability to detect the statistical structure of stimuli and monitor others' visual perception allows infants to form an implicit understanding of others' behavior. Over time, this implicit understanding may develop into an explicit Theory of Mind through executive function and language-based social and communicative interactions (Devine & Hughes, 2014; Symeonidou et al., 2024; Tomasello, 2018).

The shared intentionality theory emphasizes social and communicative interactions with others as the primary drivers of developmental change of false belief understanding (Tomasello, 2018). According to this theory, the evolution of explicit false belief understanding begins with the coordination of perspectives during joint attention in infancy, leading to an explicit understanding of belief by preschool age (Tomasello, 2018). Conventional language facilitates the public expression of mental content, making it accessible for joint attention (Schüler et al., 2024b). Through joint attention, individuals can appreciate differing perspectives. When discussing truth-bearing propositions, conflicting perspectives may arise, which do not always align with objective reality. As children develop an understanding of objective perspectives, such discourse helps them distinguish between subjective views (e.g., appearance, belief) and objective facts (e.g., truth, reality). For example, the TOMII/TOMEC study found that declarative joint attentional skills at 12 months—where infants pointed to an object outside the experimenter's view—predicted false belief understanding at 50 months, irrespective of imperative pointing or broader cognitive functioning (Sodian et al., 2020). This specificity was further supported by neural evidence (Kühn-Popp et al., 2016).

In addition, the false belief understanding observed in infants and toddlers gradually develops into its adult form, potentially through conceptual enrichment and enhanced integration with executive control (Carruthers, 2013, 2016). Recent research supports this perspective. In a study by Sodian et al. (2024), performance on a specialized location false belief task—focusing on the causal relationship between beliefs and action—was predicted solely by executive function, suggests that developmental progress in executive functions plays a crucial role in forming a robust understanding of the causal link between beliefs and action. This interpretation aligns with the developmental enrichment perspective, indicating that changes in executive function contribute incrementally rather than fundamentally to false belief understanding.

1.2.3. Evidence of longitudinal relationships between implicit and explicit false belief understanding

The implicit false belief paradigm reveals early false belief understanding in infants and toddlers through cross-sectional analyses of spontaneous response. Based on the emergence of implicit false belief understanding, the developmental trajectory suggests that implicit understanding precedes explicit understanding. However, the relationship between implicit and explicit false belief understanding can only be comprehensively examined through longitudinal data, which accurately reflects the developmental sequence. Furthermore, longitudinal studies tracking the development of false belief understanding from infancy to early childhood provide crucial evidence for theoretical accounts of early and later false belief understanding. Although rare, some studies assessed children's implicit and explicit false belief understanding at least twice over intervals of months or years (Kloo et al., 2020, 2022; Low, 2010; Poulin-Dubois, 2020; Poulin-Dubois et al., 2023; Sodian et al., 2016, 2024; Thoermer et al., 2012). These studies provide valuable longitudinal evidence, with some revealing correlations between implicit and explicit false belief understanding, though other studies did not find these results.

In a comprehensive longitudinal study, researchers assessed both implicit and explicit false belief understanding, along with executive function and other general cognitive abilities, from infancy through age 6 (see Sodian et al., 2020, for a review). Three studies that utilized data from this longitudinal sample reported consistent findings regarding the early Theory of Mind (Kloo et al., 2020, 2021; Sodian et al., 2016; Thoermer et al., 2012). For example, belief-based anticipatory looking at 18 months significantly predicted verbal false belief reasoning at 48 months, independent of verbal IQ (Thoermer et al., 2012). Moreover, implicit false belief understanding at 18 months was correlated with explicit understanding between ages 4 and 5 (Kloo et al., 2020), whereas explicit false belief understanding—but not implicit understanding—was associated with executive function (Kloo et al., 2020, 2022). Additionally, in a morally relevant context, implicit false belief understanding at 18 months predicted children's comprehension of the moral intentions of an accidental transgressor at age 5, controlling for both gender and verbal IQ (Sodian et al., 2016). These findings support the idea that false belief understanding is developmentally continuous with a corresponding conceptual system in early childhood.

Recently, an independent longitudinal study[2] by Sodian et al. (2024) demonstrated that performance on a low-demands false belief task at 33 months was significantly correlated with performance on a content false belief task, as well as with a combined score of location and content false belief tasks at 52 months. This correlation remained significant even after controlling for both language ability and executive function. Additionally, a nonlinear logistic regression analysis examining location false belief performance, with predictors including the low-demands false belief task, language ability, and executive function, revealed a significant effect of executive function on location false belief performance. Furthermore, researchers also found significant correlations between early perspective-taking abilities at 27 months of age and later false belief understanding and cognitive flexibility at 52 months of age, highlighting the importance of early perspective-taking abilities for subsequent false belief understanding and flexible cognition (Kloo et al., 2024). These findings support the perspective that explicit false belief understanding develops continuously from toddlerhood into childhood. Additionally, the results suggest that executive function and early forms of perspective-taking play a crucial role in children's explicit false belief understanding, potentially supporting both moderate dual-system theory and developmental enrichment accounts.

In a short-term longitudinal study in another laboratory (Low, 2010), 102 children aged 3 and 4 completed an anticipatory looking task to assess implicit

[2] "The role of language in early Theory of Mind development," Crossing the Borders, https://crossing-project.de/, see Kaltefleiter et al., 2021, 2022).

false belief understanding, as well as content and location false belief tasks to evaluate explicit false belief understanding. A significant correlation between implicit and explicit false belief competence was found, independent of age, alongside verbal and nonverbal abilities.

In contrast, two longitudinal studies investigated Theory of Mind during infancy and preschool years (Poulin-Dubois et al., 2020, 2023). In infancy, Study 1 of Poulin-Dubois et al. (2020) used a violation-of-expectation task, whereas Study 2 of Poulin-Dubois et al. (2020) paper and Study 1 of Poulin-Dubois et al. (2023) used interactive implicit false belief tasks. Neither one of these infant false belief studies replicated the original findings. They also did not find evidence of longitudinal correlations between 14- and 18-month-olds' performance in the violation-of-expectation false belief task and later explicit false belief comprehension at ages 4 to 5, but one significant correlation of false belief understanding in an interactive helping task in infancy and an understanding of diverse beliefs at ages 4–5. However, these authors could not exclude the possibility that infants' responses to false belief tasks were due to chance, and comprehension difficulties with false belief tasks further impacted the studies during the preschool years (Sodian, 2023).

Thus, the findings on the longitudinal relationship between early and later false belief understanding abilities present a mixed and inconclusive pattern. Therefore, further longitudinal research is necessary, employing multiple methods to assess both implicit and explicit false belief understanding during infancy.

1.2.4. Relations between false belief understanding and other cognitive abilities

Developments in other cognitive domains may provide insights into the cognitive processes underlying both implicit and explicit false belief understanding, as well as their interrelationships. Empirical evidence indicates a strong association between explicit false belief understanding and other cognitive abilities, particularly executive function (Carlson et al., 2002; Devine & Hughes, 2014) and language abilities (Astington & Baird, 2005; de Villiers & de Villiers, 2014; Milligan et al., 2007), suggesting that these cognitive abilities play a crucial role in the development of false belief understanding.

Executive function

Executive function refers to a set of neurocognitive skills involved in goal-directed problem solving, including inhibitory control, working memory, and cognitive flexibility (Carlson et al., 2013a, 2013b; Sodian & Hülsken, 2014). The proximal mechanisms through which executive function contributes to the development of false belief understanding vary depending on the theoretical perspective. According to the conceptual continuity account (Baillargeon et al., 2010), traditional explicit false belief tasks require inhibitory control to suppress the tendency to respond based on one's own knowledge, as well as response-selection skills to choose the correct answer from two options. In contrast, implicit false belief tasks place minimal, if any, demands on executive function. Therefore, the observed significant correlation between executive function and false belief understanding may simply reflect superficial characteristics of the explicit false belief tasks rather than underlying conceptual processes (Hughes, 1998). That is, the executive function might enable the *expression* of a latent false belief understanding; that is, a system for understanding mental states is in place, but executive control over responses is needed for children to show what they know (Kloo et al., 2020; Leslie & Polizzi, 1998; Moses, 2001). According to the expression account, executive demands may hinder young children from effectively demonstrating their understanding of others' minds. Supporting this view, Setoh et al. (2016) found that when inhibitory-control and response-generation demands were reduced by proper experiment design (see Section 1.2.1.1), 2.5-year-old children performed reliably above chance on the standard false belief test. Additionally, a recent study by Sodian et al. (2024) indicated that the longitudinal predictive relationship between performance on the low-demands false belief task at 33 months and performance on high-demand false belief tasks at 4 years—after controlling for executive function and language—further supports the expression account.

In addition to the conceptual continuity and expression account, the relationship between executive function and false belief understanding aligns with the moderate emergence and developmental enrichment accounts, which posit that executive function is essential for the developmental emergence of false belief understanding. For example, Kloo et al. (2020) found intrinsic developmental links between implicit and explicit false belief understanding in children aged 18 to 70 months, independent of verbal ability and executive function, and observed that only explicit false belief understanding was significantly correlated with verbal IQ and executive function. These findings support theories that implicit false belief understanding evolves into explicit false belief reasoning through language and executive control development (Kloo et al., 2020). Further, Sodian et al. (2024) found that performance on a specialized false belief task focusing on the causal impact of beliefs on actions was predicted solely by executive function. Another longitudinal study on the relationship between executive function and explicit false belief reasoning found that early executive function predicts later false belief task performance but not the reverse (Carlson et al., 2004; Marcovitch et al., 2015).

Language

The emergence account posits that the development of general language competence—including syntactic, semantic, and pragmatic capacities—is essential for behavioral Theory of Mind competence in both neurotypical and neuroatypical children (Astington & Baird, 2005; Harris et al., 2005). Supporting this view, a meta-analysis by Milligan et al. (2007)

demonstrated that language abilities in neurotypical children aged 3 to 6 predict explicit false belief understanding, even after controlling for age, whereas the reverse relationship does not hold. Furthermore, well-controlled training studies involving language-based interventions for children aged 3 to 5 showed significant improvements in false belief understanding (Gola, 2012; Hofmann et al., 2016). In recent years, an ongoing debate has focused on which aspects of language are most critical for children's acquisition of false belief understanding. Notably, it has been hypothesized that mastery of complement syntax is closely related to explicit false belief understanding (de Villiers & Pyers, 2002; Grosse Wiesmann et al., 2017a; Kaltefleiter et al., 2021). Complement syntax involves replacing the object in a sentence with a subordinate clause, creating a structure in which the sentence as a whole can be true even if the main clause is false—specifically, the linguistic construction needed to express a false belief (e.g., "Leo thinks that the ball is in the box"). Empirical studies support this hypothesis. For example, research found that children's sensitivity to complement syntax is correlated with their early explicit false belief understanding (Grosse Wiesmann et al., 2017a; Kaltefleiter et al., 2021; Low, 2010), whereas implicit false belief tasks appear less dependent on language (Kulke & Rakoczy, 2019). However, contrasting evidence is provided by a study conducted by Meristo et al. (2012), which found that deaf infants of hearing parents performed significantly worse on an anticipatory looking false belief task compared to hearing children. Furthermore, Pyers and Senghas (2009) reported an extreme case involving deaf children who grew up with limited or no exposure to conventional sign language; these individuals failed nonlinguistically administered false belief tasks even in adulthood. These findings suggest that early language input is important not only for explicit but also for early non-verbal implicit false belief understanding, lending support to both the emergence account and the developmental enrichment account of early false belief understanding. However, the specific mechanisms by which language contributes to explicit and implicit false belief understanding remain to be clarified.

1.3. False belief understanding in the brain

False belief understanding is used as the prototypical problem for Theory of Mind (Fletcher et al., 1995; Gallagher et al., 2000). Research on false belief understanding has become a fascinating area of focus for both behavioral and neuroscientific studies. While earlier approaches to exploring false belief understanding were primarily motivated by cognitive science theories, recent approaches sought a neuroscientific base to enrich and expand these cognitive theories. In the last two decades, we learned a lot about the neural underpinnings of false belief understanding (e.g., Mahy et al., 2014; Mars et al., 2012a; Schneider et al., 2017; Schurz et al., 2014, 2021; Wellman, 2014). These neural findings provide valuable insights that can inform and refine cognitive theories.

Since the 1990s, researchers attempted to identify neural correlates of mentalizing processes in a broader context by comparing mentalizing tasks with physical conditions (e.g., Fletcher et al., 1995; Goel et al., 1995). For instance, Fletcher et al. (1995) conducted a study in which adult participants underwent positron emission tomography (PET) scans while reading stories that required either mental state reasoning or physical state reasoning. In a parallel study conducted the same year, Goel et al. (1995) examined participants as they engaged in a different type of mental state reasoning. They inferred functions based on the forms of both familiar and unfamiliar objects while considering the knowledge and rationality of another mind regarding these objects' functions. Both studies revealed a distinct set of brain regions, particularly the medial prefrontal cortex (MPFC), posterior superior temporal sulcus (STS), and posterior cingulate cortex (PCC), which were activated when participants reflected on others' mental states (Fletcher et al., 1995; Goel et al., 1995).

A body of cumulative research has sought to explain the mechanisms underlying false belief understanding, which is assumed to be an acid test of Theory of Mind (e.g., Saxe, 2009; Saxe & Kanwisher, 2003; Schurz et al., 2014, 2021; Schuwerk et al., 2014). Since 2000, neuroimaging studies focusing on belief reasoning frequently employed false belief stories or cartoons, contrasting them with sequences of events that do not involve mentalizing processes, such as true-belief scenarios, false-photograph tasks, or physical and mechanical reasoning (Grèzes et al., 2004; Saxe et al., 2004a; Saxe & Kanwisher, 2003; Sommer et al., 2007, 2010; Vogeley et al., 2001). The functional neuroimaging studies with neurotypical adults and older children converged to identify specific regions of the brain and cognitive processes associated with false belief understanding (Amodio & Frith, 2006; Carrington & Bailey, 2008; C. D. Frith & Frith, 1999, 2006; Gallagher & Frith, 2003; Gobbini et al., 2007; Saxe, 2009; Saxe et al., 2004a, 2004b; Schurz et al., 2014, 2021; Van Overwalle, 2009).

Schurz et al. (2014, 2021) conducted a meta-analysis of neuroimaging studies on Theory of Mind, categorizing tasks based on stimuli and instructions. Across all task types, consistent activation was observed in the MPFC and the temporoparietal junction (TPJ). These areas significantly overlap with regions implicated in mental state attribution, self- and other-related processing, and socio-affective functions (Gweon & Saxe, 2013; Molenberghs et al., 2016; Schurz et al., 2014). Moreover, research showed that regions of the default mode network[3] are also engaged during mental state reasoning tasks (Amft et al., 2015; Mars et al., 2012b; Schilbach et al., 2012). This evidence supports the concept of a core network for Theory of Mind, which is consistently activated

[3] The default mode network is often identified in studies examining spontaneous brain activity during the resting state (e.g., M. D. Fox et al., 2005; Raichle et al., 2001).

when reasoning about mental states, regardless of task or stimulus format (Leslie et al., 2004; Leslie & Thaiss, 1992). Empirical studies over the past two decades further showed that in neurotypical adults and older children, false belief understanding is supported by a mentalizing network that includes the MPFC (e.g., Sommer et al., 2010; Tholen et al., 2020; Young & Saxe, 2009) and TPJ (e.g., Grosse Wiesmann et al., 2020; Özdem et al., 2017; Perner et al., 2006), both of which are integral to the neural circuitry for false belief reasoning (see **Figure 5** and **Table 3**). The following sections will review the specific contributions of frontal and temporoparietal activity to understanding false beliefs.

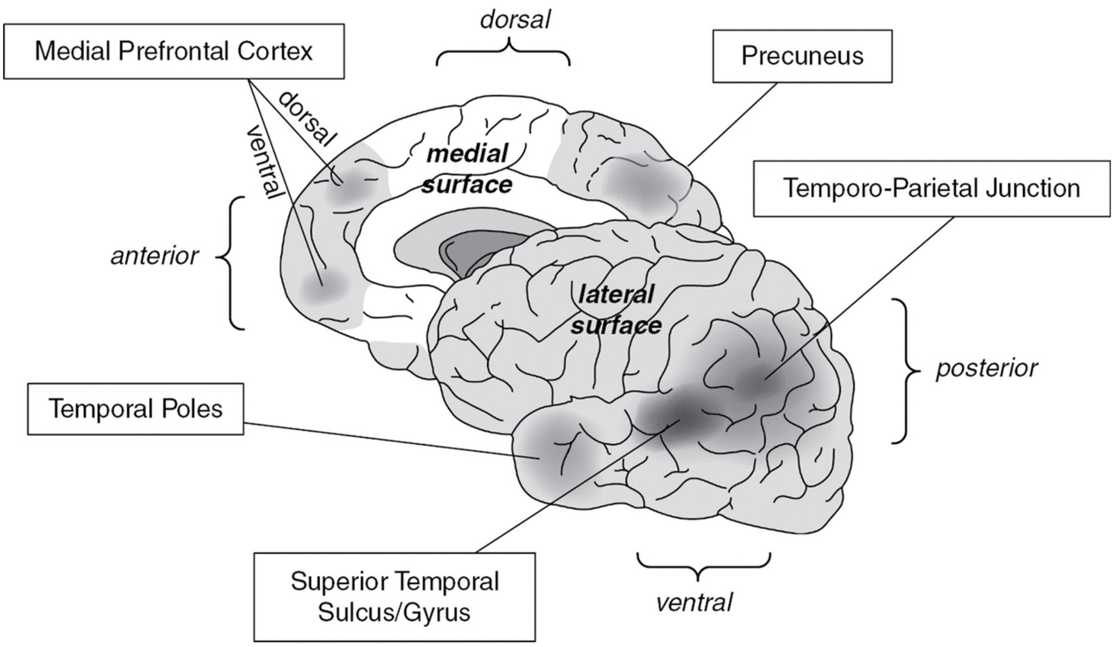


Figure 5. Depiction of the neural regions comprising the Theory of Mind network (Wellman, 2014, p. 229).

1.3.1. The role of frontal activity in false belief understanding

Numerous studies reported brain activity associated with false belief understanding in the MPFC and adjacent areas (e.g., Amodio & Frith, 2006; Stuss, 2001). These findings emphasize the involvement of frontal regions when participants reason about others' beliefs that differ from reality. The MPFC may contribute to false belief understanding through various mechanisms, including decoupling processes, the mirror neuron system, and the default mode network. Furthermore, studies of pathologies may offer additional insights into the underlying processes involved.

The MPFC is hypothesized to be involved in decoupling mechanisms, which were originally introduced into Theory of Mind research by Leslie (1987) and involve processing a mental state independently of the actual state of reality (Gallagher & Frith, 2003) (see Section 1.1.2 “Modular theory” for more details). One approach to exploring neural activity associated with the decoupling mechanism involves contrasting false belief reasoning with true belief reasoning, as false beliefs require reasoning about mental states independent of actual reality, unlike true beliefs. For instance, in the study by Grèzes et al. (2004), when participants judged actions that reflected a false belief, activation was observed in the medial prefrontal cortex (MPFC), suggesting this activation represents a mismatch between perceived actions and predicted outcomes derived from simulation. Similarly, Sommer et al. (2007) reported MPFC activity during false belief reasoning using the “Sally Anne Scenario.” Although Sommer et al. (2007) found MPFC activity linked to false belief processing, they suggested this activity might be related to both monitoring the agent’s actions (Amodio & Frith, 2006) and processing the child’s mental representation independently of stimuli (Gilbert et al., 2006). Additionally, Van der Meer et al. (2011) found that both self-perspective taking and motor response inhibition elicited MPFC activation, suggesting that the MPFC may serve as a shared neural mechanism for inhibition in both complex social contexts and basic motor responses.

Moreover, the role of frontal sites in false belief understanding may involve the human “mirror-neuron system,” a homolog primarily located in the premotor cortex, inferior frontal gyrus (IFG), and inferior parietal lobule (IPL) (Fogassi et al., 2005; Iacoboni, 2005; Rizzolatti & Craighero, 2004). This system is potentially involved in understanding actions and associated with mental states by activating one's own mirror-neuron system to simulate the observed actions and facial expressions. Some researchers proposed that the frontal sites in the mirror-neuron system support Theory of Mind. However, others argue that the mirror-neuron system is responsible for observing and matching actions, possibly for imitation, without directly inferring others' mental states (Meltzoff & Decety, 2003).

The critical role of the frontal area in false belief understanding is further supported by the studies of pathologies. Indeed, these studies indicate that frontal lobe damage, including damage to the dorsolateral frontal cortex and right ventral MPFC, is associated with impairments in false belief understanding (Happé et al., 2001; Rowe et al., 2001; Shamay-Tsoory et al., 2005; Stuss, 2001). For example, findings from Shamay-Tsoory et al. (2005) indicate that lesions in the right ventral MPFC are associated with the most pronounced deficits in Theory of Mind, highlighting the critical role of the right ventral MPFC in integrating various processes underlying Theory of Mind, thereby facilitating the development of affective Theory of Mind.

1.3.2. The role of temporoparietal activity in false belief understanding

In addition to the role of MPFC activity in false belief understanding, the TPJ has also been identified as one of the core areas in the mentalizing network, which is involved in processing social information (e.g., Frith & Frith, 2006; Samson et al., 2004; Saxe & Kanwisher, 2003). The TPJ is generally defined as an area at the border between the temporal and parietal lobes surrounding the ends of the Sylvian fissure (Schurz et al., 2014). Several neuroimaging studies reported heightened activation in the TPJ when participants engage with verbal narratives or pictorial cartoons that require inferences about a character's false beliefs, compared to physical control stimuli (e.g., Aichhorn et al., 2006; Heil et al., 2019; Özdem et al., 2017; Perner et al., 2006; Saxe & Kanwisher, 2003).

The role of the TPJ within the mentalizing network remains a topic of ongoing debate, despite consistent findings of its activation in studies investigating false belief reasoning. Some authors suggest that the TPJ is involved primarily in the perception-based processing of social information that supports false belief reasoning rather than being directly involved in the reasoning process itself (Gallagher & Frith, 2003; Hillebrandt et al., 2014; Moessnang et al., 2017). For example, using fMRI data from spontaneous mentalizing tasks with animated triangles, Hillebrandt et al. (2014) found that when movements elicited intentionality, connectivity increased between V5—a region responsible for motion detection—and the pSTS within the TPJ, which processes biological motion. Similarly, Moessnang et al. (2017) observed comparable pSTS-visual area coupling using the same task, supporting the view that the TPJ/pSTS receives signals from low-level sensory regions. However, two studies involving patients with lesions in the TPJ provided evidence against this suggestion. Apperly et al. (2004) and Samson et al. (2004) reported that damage to the TPJ leads to selective deficits in false belief reasoning, while other cognitive and executive domains remain intact. Furthermore, in neurotypical individuals, Young et al. (2010) demonstrated that transcranial magnetic stimulation (TMS) of the right TPJ affects belief reasoning. After TMS, participants showed an impaired ability to consider a person’s intent to harm when making moral judgments about acts of violence. Moreover, some suggested that the TPJ is specifically involved in thinking about another person’s thoughts (e.g., Perner et al., 2006; Saxe & Powell, 2006), although the area has also been found to be active during more general attention-related tasks, which has inspired alternative theories about its functions. For instance, Perner et al. (2006) found that the right TPJ is involved in computing mental states that create a perspective difference, such as a person's false belief that contrasts with the state of reality. Sommer et al. (2007) also observed differential brain activity for false over true belief reasoning, thus arguing a specific right TPJ role in belief reasoning. The activation of the TPJ may be related to the computation of mental representations that create perspective differences, such as a person's false belief that contrasts with reality, suggesting a central role in the decoupling mechanism. On the other hand, the right TPJ is also argued to be involved in general belief reasoning, encompassing both true and false beliefs (Saxe, 2009; Saxe & Kanwisher, 2003). By contrasting true and false belief reasoning with a non-mental control task (false photograph task), Aichhorn et al. (2009) observed right TPJ activity for both belief tasks, supporting the assumption of right TPJ involvement in general belief reasoning.

In addition, it has been proposed that distinct clusters exist within the TPJ, with one cluster being involved in attentional processes and the other in social cognition (Krall et al., 2015). The TPJ appears to be primarily implicated in the reorientation of attention and is also activated during tasks involving social cognition, as such tasks inherently require attentional reorientation along with other domain-general processes (Dugué et al., 2018; Geng & Vossel, 2013). For example, Krall et al. (2015) demonstrated that continuous theta burst stimulation applied to the TPJ disrupted both attentional control in a spatial cueing paradigm and mentalizing in a false belief task, with a positive correlation between the effects on both tasks. Schuwerk et al. (2017, 2021) adapted a spatial cueing paradigm to make another person's mental state task-relevant by manipulating participants' beliefs about the cue's origin. They identified two distinct subregions within the right TPJ—the STS and the SMG—both of which were responsive to attentional cueing and social context manipulation. Gallagher & Frith (2003) suggest that the STS may be involved in false belief understanding because it is associated with more general processing of biological and autobiographical memory retrieval. These findings suggest that the right TPJ may be involved in both attentional control and Theory of Mind networks, with its function shaped by context-dependent coupling with the respective network.

Interim summary on the role of frontal and temporoparietal activity in Theory of Mind: In adults, the MPFC and TPJ are primarily functionally associated with false belief reasoning and show strong preferential activation when individuals consider the mental states of others. However, findings from adult studies primarily focus on the neural mechanisms underlying Theory of Mind in fully developed individuals and cannot elucidate the brain and cognitive processes involved at earlier developmental stages (Wellman, 2018). Thus, prior neuroimaging studies in adults leave critical questions unanswered regarding the early development of Theory of Mind. Over the past two decades, methodological advancements have enabled researchers to directly assess activity in brain regions associated with Theory of Mind during early childhood and infancy (e.g., Quesque et al., 2024; Schurz et al., 2021; Schuwerk, Kampis, et al., 2021). These studies started to bridge the gap between the extensive knowledge of early behavioral development in false belief understanding and the mature neural profile underlying false belief reasoning. Furthermore, they offer accumulating evidence that mentalizing brain regions operate as an integrated network, displaying preferential responses for processing mental states from an early stage of development (Richardson & Saxe, 2020a, 2020b).

1.3.3. False belief understanding in children's brain

A fundamental gap in previous empirical research and theoretical discussions on the neural foundations of false belief understanding is the scarcity of developmental data and limited consideration of developmental trajectories. The majority of neuroimaging studies focused on autistic populations and neurotypical adult subjects, using structural and functional MRI. However, data concerning the development of the neurotypical brain in relation to false belief understanding remain notably sparse (Beaudoin et al., 2020; S. Li et al., 2023; Rakoczy, 2022). To date, most neuroimaging research on the development of false belief understanding in children has primarily targeted older children, aged 6–12 years, an age range when this cognitive ability is already well-established (e.g., Gweon et al., 2012; Kobayashi et al., 2007a, 2007b; Saxe et al., 2009; Sommer et al., 2010). To address critical unresolved questions about the neural development of false belief understanding, researchers must conduct studies assessing mentalizing network activity in children under the age of 6 (Grosse Wiesmann et al., 2020, 2017b; Hyde et al., 2018; Moraczewski et al., 2018; Richardson et al., 2018). Although some initial studies have begun to explore this area, the existing evidence concerning very young children remains insufficient (Richardson & Saxe, 2020a).

Several initial studies have found that children aged 6 to 12 years, like adults, preferentially recruit the bilateral MPFC and TPJ when reasoning about the minds of others (e.g., Gweon et al., 2012; Kobayashi et al., 2007a, 2007b; Saxe et al., 2009; Sommer et al., 2010). For example, Kobayashi et al. (2007a) presented verbal and non-verbal second-order false belief tasks to children aged 8 to 12 years and to adults. Both age groups exhibited significant activity in the TPJ and the right IPL in a modality-independent manner, suggesting that these regions play a crucial role in Theory of Mind across both childhood and adulthood, regardless of the modality (Kobayashi et al., 2007a). Furthermore, Saxe et al. (2009) and Gweon et al. (2012) found evidence of increasing response selectivity in the right TPJ for situations requiring the attribution of thoughts to other agents or when listening to descriptions of characters' mental states, compared to descriptions of physical events. This increasing selectivity continues throughout childhood (ages 6 to 11), suggesting that the TPJ, which is critical for attributing mental states, becomes progressively more efficient and specialized in its functioning during childhood. However, Sommer et al. (2010) compared false and true belief reasoning in children (ages 10-12) and adults using cartoon stories depicting unexpected transfers. Both groups showed activation in the dorsal MPFC during false belief reasoning. Unlike adults, children did not selectively recruit the right TPJ and exhibited greater activation in the right rostral prefrontal cortex and PCC. These findings suggest that the cortical network supporting false belief reasoning undergoes significant developmental changes. Nevertheless, the youngest children participating in these studies were at least 6 years old, whereas the developmental age range of explicit and implicit false belief understanding (e.g., Onishi & Baillargeon, 2005; Wellman et al., 2001; Wimmer & Perner, 1983) plausibly indicates that cortical regions supporting Theory of Mind may already be functioning in toddlers or even infants, emerging well before the ages previously studied. Further studies involving younger children would provide a clearer picture of the development of the neurobiological underpinnings of false belief understanding.

Some neuroimaging studies have explored the early sensitivity to mental states in infants and toddlers, though such studies remain limited. For example, Richardson et al. (2018) and Moraczewski et al. (2018) recorded fMRI responses in children (ages 3-6) while they viewed wordless animated films. The time course of the mentalizing network in these children showed a significant correlation with that of adults but not with other control networks. Furthermore, Hyde et al. (2018) found that even in 7-month-olds, the right TPJ showed preferential activation during false belief sequences, especially when the target person returned to retrieve an object. Notably, the response in infants was strikingly similar to that observed in adults (Hyde et al., 2015), suggesting that similar mental state processing, supported by the mentalizing network, occurs from infancy through adulthood.

Additionally, Richardson et al. (2018) found that Theory of Mind brain regions showed significantly stronger correlations in children who passed than in those who failed false belief tasks. However, these within-network correlations did not remain significant after controlling for age. It is possible that correlations within the mentalizing network are particularly important for, or indicative of, mentalizing abilities in young children. These findings are consistent with the following two studies. First, Grosse Wiesmann et al. (2017b) used diffusion tensor imaging to examine white matter connections related to Theory of Mind development in young children. They observed increased connectivity in tracts surrounding key mentalizing regions, such as the MPFC and TPJ, in 3- and 4-year-olds who experienced a developmental breakthrough in false belief understanding. Second, Xiao et al. (2019) found that resting-state connectivity within the TPJ correlated with Theory of Mind abilities, as assessed by a parent questionnaire, in children aged 4 to 8 years. These studies collectively suggest that the functional activity and structural connectivity of the mentalizing network are associated with Theory of Mind abilities from an early age, highlighting the critical role of this specialized network in the cognitive development of young children.

Besides that, to date, three neuroimaging studies have compared the implicit and explicit false belief understanding in children (Grosse Wiesmann et al., 2020; Kobayashi et al., 2007a; Schüler et al., 2024a). Explicit false belief understanding elicited more brain activity in the right MTG and left STG/STS in all participants compared to implicit false belief understanding (Kobayashi et al., 2007a), but for children, there was no significant difference in neural activity between implicit and explicit false belief understanding. However, Grosse Wiesmann et al. (2020) found some preliminary evidence of two neural systems for thinking about others' thoughts in 3- and 4-year-old children. Grosse

Wiesmann et al. (2020) linked false belief understanding in 3- to 4-year-olds to cortical surface area and thickness maturation. They found that explicit false belief reasoning was supported by cortical structures in the precuneus and TPJ, regions also involved in adult false belief understanding. In contrast, implicit false belief reasoning was linked to a subregion of the TPJ—the SMG—which is associated with emotional perspective taking, action observation, and social attention. Further, Schüler et al. (2024a) revealed a ventral network for implicit false belief understanding and a dorsal network for explicit false belief understanding. Implicit false belief understanding abilities are linked to the maturation of ventral fiber tracts in the salience network, which is involved in bottom-up social attention, suggesting that young children may adopt perceptual, social cues to predict agents' actions (Schüler et al., 2024a). Conversely, explicit false belief understanding performance correlates with the maturity of the arcuate fascicle and cingulum, which connect to the default network, engaged in higher-order social cognition. These neural dissociations between implicit and explicit false belief understanding suggest two systems for reasoning about others' minds, with explicit understanding emerging around age 4, while implicit understanding relies on earlier socio-cognitive processes. Considering the inconsistency between these studies, further evidence on the potential neural basis for the differences between implicit and explicit false belief understanding should be conducted (see Study 2 in this dissertation).

The results of neuroimaging studies have sought to "localize" the neural regions involved in false belief reasoning with high spatial resolution. However, due to the limited temporal resolution of neuroimaging techniques, these studies cannot clarify how these regions engage in the detailed, nuanced cognitive processes underlying false belief understanding. Addressing questions about the cognitive processes involved, therefore, requires EEG studies, which provide higher temporal precision (Sabbagh, 2013). Identifying an electrophysiological marker for false belief reasoning is essential for three reasons. First, ERP techniques offer a significant advantage by precisely characterizing the timing of specific neurocognitive processes, yielding insights into the task-dependent neurophysiological foundations of false belief understanding. Second, beyond task-dependent ERP, task-independent EEG methods can offer deeper insights into the potential causal relationship between baseline brain activity and the development of false belief understanding. Third, numerous theories currently seek to explain the cognitive mechanisms underlying false belief reasoning in young children (e.g., Baillargeon et al., 2016; Doherty & Perner, 2020; Sodian et al., 2020; Sodian & Kristen, 2016). Exploring potential electrophysiological markers for false belief reasoning could be instrumental in refining theoretical frameworks and guiding future research in this intriguing area.

Table 3. A summary of the results from neuroimaging studies investigating the neural correlates of false belief understanding.

ID	Research	Participants	Task type	MPFC	IFG	SMG	ACC	Motor cortex	Paracingulate	PC/PCC	TL	MTG	TPJ	IPL	Amygdala	Occipital cortex
1	(Grosse Wiesmann et al., 2020)	C	Verbal							1		1-R	1-R			
2		C	Nonverbal			1-R				1-L						
3	(Boccadoro et al., 2019)	A	Nonverbal	1-R	1-L	1-R						1-R		1		
4	(Heil et al., 2019)	A	Verbal		1-R	1-R			1-L	1-L	1		1	1		
5	(Hughes et al., 2019)	A	Verbal	1		1-R				1-R	1-R		1			1
6	(Moraczewski et al., 2018)	C	Nonverbal	1									1			
7	(Hyde et al., 2018)	C	Nonverbal										1			
8	(Oliver et al., 2018)	A	Verbal	1						1	1		1			
9	(Grosse Wiesmann et al., 2017b)	C	Nonverbal	1-R						1-R		1-L	1-R			
10	(Naughtin et al., 2017)	A	Nonverbal	1-L						1	1-R		1-R			
11	(Özdem et al., 2017)	A	Nonverbal							1			1-R			
12		A	Verbal	1-R								1-R				
13	(S. M. Lee & McCarthy, 2016)	A	Verbal	1	1					1	1		1			1
14	(Hartwright et al., 2015)	A	Verbal	1						1	1	1	1			1-L
15	(Andrews-Hanna et al., 2014)	A	Verbal	1						1	1		1		1	
16	(Corradi-Dell'Acqua et al., 2014)	A	Verbal	1	1					1	1	1	1			
17	(Contreras et al., 2013)	A	Verbal	1							1		1			
18	(Döhnel et al., 2012)	A	Nonverbal	1	1-L			1-R		1			1			
19	(Dodell-Feder et al., 2011)	A	Verbal	1						1	1-R		1			
20	(J. Lee et al., 2011)	A	Verbal	1						1-R	1-R		1			
21	(Young et al., 2011)	A	Verbal	1						1			1			
22	(Van der Meer et al., 2011)	A	Verbal	1-L	1					1		1				

23	(Sommer et al., 2010)	A & C	Nonverbal	1	1-R		1			1				1		
24	(Young, Dodell-Feder, et al., 2010)	A	Verbal	1						1			1			
25	(Saxe et al., 2009)	C	Verbal	1						1			1			
26	(Aichhorn et al., 2009)	A	Verbal	1	1-L					1	1-L	1	1			
27	(Young & Saxe, 2009)	A	Verbal	1						1	1-R		1			
28	(Kliemann et al., 2008)	A	Verbal	1						1			1			
29	(J. P. Mitchell, 2008)	A	Verbal	1		1				1			1-R			
30	(Gobbini et al., 2007)	A	Verbal	1	1	1	1	1		1	1	1	1	1		
31	(Kobayashi et al., 2007a)	A & C	Nonverbal	1-R									1	1-R		1
32	(Sommer et al., 2007)	A	Nonverbal	1-R			1			1		1-R	1-R	1-R		
33	(Young et al., 2007)	A	Verbal	1						1	1-R		1-R			
34	(Perner et al., 2006)	A	Verbal		1-R	1-L						1	1			
35	(Saxe et al., 2006)	A	Verbal	1						1	1-R		1		1	
36	(Saxe & Powell, 2006)	A	Verbal	1						1	1		1		1-L	
37	(Saxe & Wexler, 2005)	A	Verbal	1						1			1-L			
38	(Grèzes et al., 2004)	A	Nonverbal	1					1		1	1				1
39	(Saxe & Kanwisher, 2003)	A	Verbal	1						1	1		1			
40	(Gallagher et al., 2000)	A	Verbal	1							1		1			
41		A	Nonverbal	1						1			1-R			
	Total (41)			35	10	7	3	2	2	32	20	12	35	6	3	5
	Percent			85	24	17	7	5	5	78	49	29	85	15	7	12

Note: A = adults; C = children. MPFC = medial prefrontal cortex; IFG = inferior frontal gyrus; ACC = anterior cingulate cortex; PC = precuneus; PCC = posterior cingulate cortex; TL = temporal lobes; SMG = supramarginal gyrus; MTG = middle temporal gyrus; TPJ = temporoparietal gyrus; IPL = inferior parietal lobe. L = left hemisphere; R = right hemisphere.

1.4. Event-Related Potential (ERP) studies of belief reasoning

Over the last two decades, significant progress has been made in identifying the neural correlates of false belief understanding. One way to partially resolve the nuanced cognitive processes involved in false belief understanding is to characterize the "mentalizing network" temporally using ERP. ERP is widely used as a relatively low-cost (per subject), non-invasive technique in cognitive neuroscience that makes few physical demands of study participants (Boustani, 2023). Unlike MRI techniques, ERP has a high temporal resolution, allowing the millisecond-by-millisecond visualization of brain activity (Sabbagh, 2013). Therefore, ERP would be more sensitive to temporal variations that may be associated with detailed cognitive processes in false belief understanding. ERP signals from individual trials are averaged to extract the stable and reliable characteristics of brain responses associated with false belief understanding. In recent years, the widespread use of ERP methodology has enabled researchers to investigate this question from a cognitive neuroscience perspective, leveraging its high temporal resolution (Sabbagh, 2013).

1.4.1. ERP paradigms in exploring the neural basis of belief reasoning

The acquisition of false belief reasoning is considered to be a key entailment in Theory of Mind development. The electrophysiological mechanisms involved in false belief reasoning have been explored by using ERP methodology (Sabbagh, 2013). As Theory of Mind representation requires the establishment of a situational context and a narrative to convey the knowledge of the agents' world, a considerable part of the ERP research on false belief has focused on implementing a compatible version of the classic unexpected transfer task (i.e., Max's chocolate story, Wimmer & Perner, 1983). In terms of paradigm design, previous neurophysiological studies can be broadly categorized into two groups based on whether belief reasoning was compared with another mental or non-mental state: belief reasoning vs. non-mental state processes (e.g., false photo, false sign, and reality understanding), and false belief reasoning vs. other mental states reasoning (e.g., true belief, desire, and pretense) (see **Table 4**).

The first type of research compares neural activity associated with belief reasoning to that elicited by non-mental representations, such as reality understanding, false photographs, or false signs. For example, Liu et al. (2004, 2009b) compared ERPs elicited by judgments based on belief reasoning to those based on reality for adults and children. The results demonstrated that a late waveform peaking around 800 ms post-stimulus—with a frontal scalp distribution consistent with a possible source in the orbitofrontal cortex—differentiated judgments about belief from those about reality. Another closely matched control condition for the processing demands of the standard false belief task, for example, is the “false photograph” or ‘‘false sign’’ task (Perner & Leekam, 2008). The “false photograph” task requires participants to answer questions about a past physical representation (i.e., a photograph) that was once true but is now false. Similarly, the “false sign” task involves interpreting a symbolic representation that misrepresents the current reality (e.g., a sign indicating an object is in location A while the object has moved to location B). Both tasks serve as tests of domain specificity for brain regions, as they share similar cognitive demands with the false belief task but differ in their requirement for reasoning about beliefs (Gweon & Saxe, 2013). Any differences in neural responses between false belief tasks and false photograph or false sign tasks reflect the specific need to consider mental states rather than differences in inhibitory or executive demands (Saxe et al., 2009). For instance, Sabbagh and Taylor (2000) asked participants to reason belief representations and photograph representations. Results showed that neural activity elicited by mental representations (beliefs), as compared to non-mental representations (photographs), was characterized by increased focal positivity over frontal areas and decreased positivity over parietal areas. In a location transfer false-sign task, an arrow initially points to location A, where the target object is located; however, after the object moves to location B, the arrow continues to indicate the original location. Participants are asked to identify the object's location based on the sign (Sabbagh et al., 2006; Zhang et al., 2013). Findings revealed that the late slow waveform for false belief was more positive at frontal, central, and parietal regions compared to the false sign condition.

The second kind of research mainly focuses on ERP differences between false belief reasoning and other types of mental states, such as desire reasoning, pretense, etc. For instance, false belief reasoning differs from true belief reasoning (e.g., Chen et al., 2012; Geangu et al., 2013; Meinhardt et al., 2011; Zhang et al., 2013), desire reasoning (Bowman et al., 2012; Liu et al., 2009a), and pretense (Kühn-Popp et al., 2013; Meinhardt et al., 2012). Bowman et al. (2012) and Liu et al. (2009a) investigated the neural correlates of belief reasoning, desire reasoning, and a physical control condition in adults and children. They found that mid-frontal scalp activations were associated with both belief and desire reasoning, but analyses using only correct trials revealed selective right-posterior activations specific to belief reasoning. Kühn-Popp et al. (2013) and Meinhardt et al. (2012) explored the neural basis of false belief reasoning, pretense, and reality, which are assumed to reflect the decoupling mechanism of meta-representation. Results revealed late anterior activation (600–900 ms) for false belief reasoning, likely reflecting the decoupling mechanism in meta-representation. Temporal and topographic differences between false belief and pretense suggest distinct neural substrates, challenging meta-representational interpretations of pretense (Kühn-Popp et al., 2013; Meinhardt et al., 2012).

Besides, Directly comparing false belief and true belief conditions is an effective method to investigate the specificity of false belief reasoning by contrasting scenarios that capture the core of false belief understanding (Meinhardt et al., 2011). This comparison involves assessing a condition where an individual's mental state is independent of reality (false belief) against one where it aligns with reality (true belief) (Perner, 1991). The advantage of this comparison lies in the fact that both conditions require action prediction or explanation. Meinhardt et al.'s (2011) study narrated a story via an ERP paradigm according to the "Sally Anne Scenario" (Baron-Cohen et al., 1985) in which participants watched a scene in which an agent forms a mental state (such as a true or false belief); the participant was shown the agent acting in accordance with their belief (expected) or not acting in that way (unexpected). In the task, participants viewed a sequence of images depicting two children (e.g., Betty and Nick) and two boxes. Betty placed an object (e.g., a teddy bear) into a box and then left the room. Nick subsequently moved the teddy bear to the other box. In the true belief condition, Betty returned and observed this action, whereas in the false belief condition, Betty was absent when Nick moved the teddy bear. The final image showed Betty searching for the teddy bear, and participants were asked whether they thought Betty would search in that location. Results identified two waveforms distinguishing the false belief from the true belief conditions: a late positive complex, associated with reorienting attention from external stimuli to internal mental representations, and a late anterior slow wave, linked to stimulus-independent processing of mental representations, potentially central to the decoupling mechanism (Meinhardt et al., 2011). In another ERP study, Geangu et al. (2013) developed a simplified non-verbal false belief task similar to that used by Onishi & Baillargeon (2005). Sequences of still photographs depicted events inducing either a true or false belief in the protagonist. In both conditions, the protagonist placed an object in one of two boxes, and her view was obstructed by an occluder. The object either remained in the original box (true belief) or was moved to the other box (false belief). When the occluder was removed, the protagonist reached for the object, reflecting either a true or false belief about its location. Results showed that passive viewing of stimuli led to different ERP amplitudes between false and true belief conditions across frontal, central, and parietal regions. These findings suggest that (i) frontal activity is essential for processing false belief tasks, and (ii) parietal activation related to monitoring others' beliefs can occur without explicit mentalizing instructions (Geangu et al., 2013). Another ERP study by Zhang et al. (2009) compared brain activity elicited by a standard false belief task to an adapted version. In the standard task, participants observed an object being moved from location A to B, whereas in the adapted task, the object's new location was unknown. The adapted task was proposed to have lower inhibitory demands, as participants did not need to inhibit their knowledge of the object's actual location. Results showed that, in standard tasks, the late positive component elicited by false belief reasoning was more positive at 470–520 ms over the middle frontal gyrus than in adapted tasks. These findings confirm that inhibitory control influences false belief reasoning, suggesting that inhibition occurs prior to the formation of false belief concepts.

Given the accumulating evidence of neural correlates of explicit false belief understanding in adults and older children, and especially Setoh's work in designing the first behavioral explicit false belief task with fewer demands, demonstrating that toddlers younger than 3 years old can attribute explicit false belief understanding, it is imperative to devise an ERP paradigm tailored to this young age group. Consequently, we have developed an ERP belief paradigm, termed "Leo's belief task," designed to evaluate neural responses in toddlers as they process information pertaining to a protagonist's (Leo's) beliefs, whether true or false. In Study 1, our aim was to explore the neural correlates of false belief understanding in children around 33 to 36 months old.

1.4.2. *Typical waveforms associated with belief reasoning*

Evidence from multiple existing ERP studies shows that several typical ERPs are associated with false belief understanding, including the late slow waveform and late positive complex (LPC). Besides these two ERPs, recently, some researchers also indicated that the N400 has a potential correlation with false belief understanding.

Late slow waveform

Previous ERP studies regarding the late slow waveform effect (either positive or negative) associated with mental state processing in Theory of Mind have been replicated with many different kinds of stimuli (e.g., Geangu et al., 2013; Kühn-Popp et al., 2013; Liu et al., 2009b, 2009a; Meinhardt et al., 2011, 2012). Slow waveforms are tonic, with less distinct peaks, persisting from several hundred milliseconds to several seconds (Rösler et al., 1997). They display task-specific topography, amplitude, and temporal extent (Ruchkin et al., 1995). The topography varies with system demands, with maximum amplitude patterns emerging over various cortical areas. Amplitude increases with task difficulty, reflecting the cognitive effort required. The temporal extent correlates with the duration of cognitive subroutines, changing as different neural modules activate during the task. In Theory of Mind tasks, the late slow waveform typically emerges around 600 ms post-stimulus and manifests extensively across the frontal and parietal electrode sites, a pattern frequently observed when participants engage with complex visual stimuli that require the attribution of mental states to others (Keil, 2013; Sabbagh, 2013).

Late slow waveforms are differentially elicited by mental state attribution processes versus non-mental state attribution processes (e.g., Bowman et al., 2012; Guan et al., 2020; Liu et al., 2009a, 2009b; Sabbagh & Taylor, 2000). For example, false belief reasoning, compared to false photo reasoning or false sign understanding, induces a left-frontal late slow waveform from 600 to 840 ms post-stimulus (Sabbagh & Taylor, 2000) or a late slow waveform across frontal, central, and parietal sites from 700 to 880 ms post-stimulus (Zhang et al., 2013). Belief reasoning, encompassing both false belief and true belief versus reality understanding elicits a negative left-frontal late slow waveform from around 700 to 900 ms in adults (Liu et al., 2004, Liu et al., 2009b) and a frontocentral late slow waveform from 400 to 1000 ms (Chen et al., 2012), whereas in children, it appears from 1,400 to 1,500 ms (Liu et al., 2009b) and from 600 to 920 ms in parieto-occipital sites (Guan et al., 2020; Kühn-Popp et al., 2013).

Late slow waveform variations also exist across different mental state processes. For instance, the late slow waveforms for false belief reasoning differ from those of true belief reasoning (e.g., Chen et al., 2012; Geangu et al., 2013; Meinhardt et al., 2011; Zhang et al., 2013), desire reasoning (Bowman et al., 2012; Liu et al., 2009a), and pretense (Kühn-Popp et al., 2013; Meinhardt et al., 2012). Compared to true belief reasoning, false belief reasoning induces a midfrontal late slow waveform from 600 to 1000 ms in adults (Chen et al., 2012; Geangu et al., 2013; Meinhardt et al., 2011; Zhang et al., 2013) and a broad anterior component from 750 to 1450 ms in children (Meinhardt et al., 2011).

Late slow waveforms are widely observed in mentalizing ERP studies despite varying in polarity, time windows, and scalp distribution depending on the complexity of mental state reasoning (Meinhardt et al., 2011). However, their interpretation of belief reasoning remains controversial due to diverse paradigms. Late slow waveforms were proposed to reflect processes associated with decoupling mental states from reality (e.g., Liu et al., 2004) or the decoupling mechanism involved in meta-representation (Kühn-Popp et al., 2013; Meinhardt et al., 2012). Additionally, they have been linked to working memory processes (Bailey et al., 2016; Barriga-Paulino et al., 2014), suggesting they may reflect extended working memory processing and allocation of attentional resources during belief reasoning. Given the lack of consensus, further investigations into neural activities in belief reasoning may provide more evidence regarding the cognitive mechanisms underlying false belief understanding.

Late positive complex

Numerous studies employing varied paradigms also suggest the involvement of the late positive complex (LPC) in mental state reasoning, typically emerging later than 300 ms post-stimulus. For example, Cao et al. (2012) delineated LPC responses in tasks that compared simple mental state decoding, such as judging emotions from photographs, to more complex reasoning tasks, like predicting actions based on those emotions. Here, the LPC was notably evoked in reasoning scenarios, suggesting a heightened demand for integrating contextual information.

This waveform displays similarities with the P300 potential in terms of latency and posterior scalp distribution, suggesting potential shared cognitive processes (Donchin & Coles, 1988; Polich, 2007, 2012). For instance, LPC activity between 240 and 440 ms over frontal sites has been observed in tasks involving mental state reasoning and mental state decoding, compared to physical condition tasks, indicating its role in differentiating cognitive processes. Similarly, across parietal sites, the LPC distinguishes between mental state reasoning, mental state decoding, and physical conditions. Notably, in contrasting false beliefs against false photos or statements, the LPC manifests within the 300-600 ms range over both frontal and centroparietal sites, underscoring its responsiveness to nuanced cognitive distinctions in belief assessments (Sabbagh & Taylor, 2000; Y. Wang et al., 2023). Moreover, when differentiating between false and true belief reasoning, the LPC can distinctly identify these conditions within the same 300 to 600 ms timeframe post-stimulus, highlighting its potential as a neural marker of cognitive complexity in belief evaluation (Geangu et al., 2013; Jiang et al., 2016; Meinhardt et al., 2011; Y. Wang et al., 2023). The LPC shares several electrophysiological characteristics with the P300 component, particularly in scenarios necessitating the updating of a mental model. This similarity suggests that both components may participate in the processing of the mental representations updating in response to external stimuli (Donchin & Coles, 1988). For instance, a study by Y. Wang et al. (2010) demonstrated that the LPC is elicited when comprehension specifically requires Theory of Mind attributions, such as understanding mental states from cartoonish depictions of people.

The P300 is generally elicited during cognitive tasks involving attention, memory, and problem-solving, reflecting its core function in information processing and memory updating (Polich, 2007, 2012). The involvement of the P300 in Theory of Mind reasoning remains a topic of discussion. For example, research has found P300-like activity broadly focused on parietal sites in adults and exhibited a more posterior distribution in children during comparisons of false belief and true belief reasoning scenarios (Geangu et al., 2013; Meinhardt et al., 2011). Recent studies further identified a stronger posterior P300 in the false belief condition compared to the true belief condition, indicating differential cognitive processing between these belief states (Y. Wang et al., 2023). Changes in latency and amplitude during ERP tasks may reflect variations in neural processing speed and cognitive efficiency. These changes are critical in interpreting the neural underpinnings of cognitive tasks and have implications for understanding the efficiency and magnitude of neural responses in complex reasoning scenarios (Dinteren et al., 2014; Riggins & Scott, 2020).

Social N400

Mentalization, the attribution of intentions, beliefs, and desires to others, is essential for language use and comprehension. It is posited as inherently inferential, where meanings derive from semantic content and communicative intentions (Hirst, 1989). The neurocognitive underpinnings of such mentalistic inferences in linguistic contexts were explored through a series of EEG experiments focusing on the social N400 effect.

The typical N400, evoked by semantic incongruities like nonsensical combinations (e.g., “He spread the warm bread with socks”) (Kutas & Federmeier, 2011), indexes the effort involved in semantic memory retrieval and varies with the degree of semantic expectations met. In scenarios where participants and a confederate are exposed to different linguistic contexts, an N400 response emerges in participants when sentences are incongruent for the confederate but not for themselves. For example, Kampis et al. (2024) found a reduced N400 amplitude when labeling was congruent from the other’s perspective compared to incongruent, though it was always incongruent for the infant. Interestingly, this effect persists without direct instructions to consider the confederate’s comprehension, suggesting an underlying social cognitive process (Jouravlev et al., 2019; Rueschemeyer et al., 2015; Westley et al., 2017). However, the central question remains whether the social N400 primarily reflects mentalistic processes or broader social cognitive functions, such as detecting informational asymmetries. Previous studies involving infants demonstrate that the social N400 can be triggered by false belief contexts without altering semantic content, implying a sensitivity to mentalistic manipulations (Forgács et al., 2019, 2020) or altercentric bias (Kampis et al., 2024). This finding suggests that beyond mere semantic processing, the N400 might also be indicative of mentalization.

Neuroimaging links the N400 response to brain regions involved in both language processing and mentalization, such as Wernicke's area, the STS, and the TPJ, which are known for their roles in processing false beliefs (Federmeier & Laszlo, 2009; Lau et al., 2009; Van Petten & Luka, 2006). These associations point to a possible shared neural basis between semantic comprehension and the mentalistic computations required for effective communication (Schurz et al., 2021).

Table 4. A summary of the results from ERP studies investigating the neural correlates of false belief understanding.

ID	Research	Participants	Samples	Ages	ERP pre-processing re-reference	ERP trigger	Paradigm type (non-verbal vs. verbal)	Condition comparison	ERPs, Scalp distribution, Timing
1	(Y. Wang et al., 2023)	A	16	19.80 years	linked mastoid reference	After explicit instruction	Verbal	False belief vs. False statement vs. true belief	P300, centroparietal sites, 250-450ms, false belief vs. false statement/true belief
2	(Guan et al., 2020)	C	30	10.91 years	common average reference	After explicit instruction	Verbal	Belief condition (false- and true-belief reasoning) vs. Complementary condition (false- and true-complement)	Late slow waveform, parieto-occipital sites, 600-900ms.
3	(Jiang et al., 2016)	A	16	21.28 years	linked mastoid reference	After explicit instruction	Verbal	self-FB vs. self-TB other-FB vs. other-TB	LPC, frontal and frontocentral sites, 450-600ms, self-FB vs. other-FB P2, frontal and frontocentral sites, 120-200ms, FB vs. TB N2, frontal and frontocentral sites, 200-400ms, FB vs. TB
4	(Geangu et al., 2013)	A	18	22.70 years	linked mastoid reference	Passive paradigm	Non-verbal	false- vs. true-belief reasoning	Late slow waveform, frontal sites, 785-840ms; central sites, 570-690ms; parietal sites, 550-850ms. LPC, parietal sites, 185-275ms.
5	(Zhang et al., 2013)	A	14	24.30 years	linked mastoid reference	After explicit instruction	Non-verbal	false-belief vs. false-sign/true-belief reasoning	Late slow waveform, frontal, central, parietal sites, 800-880ms.
6	(Kühn-Popp et al., 2013)	C	21	7.34 years	common average reference	After explicit instruction	Verbal	false-belief reasoning vs. reality	Late slow waveform, frontocentral sites, 600-920ms; parieto-occipital sites, 420-920ms.
7	(Bowman et al., 2012)	C	18	8.11 years	common average reference	After explicit instruction	Verbal	belief reasoning vs. desire reasoning vs. physical control	Late slow waveform: Belief/Desires vs. physical: mid-frontal slow sites, 200-250ms, 350-850ms Belief vs. Desire judgment: right posterior sites, 600-800ms, 850-1400ms

8	(Meinhardt et al., 2012)	A	24	23.5 years	common average reference	After explicit instruction	Verbal	false-belief reasoning vs. reality	Late slow waveform, midfrontal sites, 600-900ms
9	(Cao et al., 2012)	A	13	20.4 years	linked mastoid reference	After explicit instruction	Non-verbal	mental-state reasoning vs. mental-state decoding vs. physical condition	N1, parietal sites, 70-110ms, reasoning vs. physical. P2, prefrontal sites, 120-160ms, reasoning vs. physical. N2, prefrontal sites, 170-240ms, reasoning vs. physical; parietal sites, 170-240ms, reasoning vs. physical. LPC, prefrontal sites, 240-440ms, reasoning vs. decoding/physical; parietal sites, 240-440ms, reasoning/decoding vs. physical
10	(Chen et al., 2012)	A	23	22.6 years	linked mastoid reference	After explicit instruction	Non-verbal	false belief vs. true belief non-factive verb comprehension vs. factive verb comprehension	Late slow waveform, fronto-central sites, 400-1000ms, false belief vs. true belief, non-factive verb comprehension vs. factive verb comprehension
11	(Meinhardt et al., 2011)	A & C	Adults: 21 Children : 22	Adults: 24.33 years Children: 7.7 years	linked mastoid reference	Passive paradigm	Non-verbal	true / false belief reasoning	LPC, 300-600ms Adults: central sites; Children: parietal sites Late slow waveform Adults: midfrontal sites, 600-900ms. Children: broad anterior sites, 750-1450ms.
12	(Liu et al., 2009b)	A & C	Adults: 24 Children : 44	Adults: did not provide Children: 5 years 11 months	common average reference	After explicit instruction	Non-verbal	Belief/Think condition vs. Reality condition	Late slow waveform Adults: left frontal sites, 775-850ms Child passers: left frontal sites (a more diffuse frontal scalp distribution), 1400-1500ms.
13	(Liu et al., 2009a)	A	15	22 years	common average reference	After explicit instruction	Verbal	belief reasoning vs. desire reasoning vs. physical control	Late slow waveform midfrontal sites, 800-850ms, belief/desire vs. physical. right posterior sites, 600-700, belief vs. physical. 700-800ms, belief vs. physical, belief vs. desire.
14	(Zhang et al., 2009)	A	14	22.5 years	common average reference	After explicit instruction	non-verbal	Standard (move to another location) vs. Adapted (move away from the scene)	LPC, anterior sites, 470-520ms.

15	(Y. Wang et al., 2008)	A	14	21.7 years	linked mastoid reference	After explicit instruction	Verbal	false belief reasoning	LNC, frontal sites, 400-800ms.
16	(Liu et al., 2004)	A	17	19-35 years	linked mastoid reference	After explicit instruction	Non-verbal	Belief/Think condition vs. Reality condition	Late slow waveform, left frontal sites, 700-900ms.
17	(Sabbagh & Taylor, 2000)	A	23	18-42 years	common average reference	After explicit instruction	Verbal	false-belief reasoning vs. false-photo	LPC, left frontal sites, 300-400ms. LNC, left parietal sites, 300-400ms. Late slow waveform, left frontal sites, 600-840ms.

Note: A = adults; C = children. LPC = late positive complex, LNC = late negative component, FB = false belief, TB = true belief.

1.4.3. *ERP correlates of belief reasoning in adults and children*

Developmental investigations comparing neural activation between 6- to 12-year-old children and adults consistently showed that children demonstrate a more diffuse distribution of ERP waveforms compared to adults (e.g., Guan et al., 2020; Kühn-Popp et al., 2013; Liu et al., 2009b; Meinhardt et al., 2011). Seven ERP studies (Bowman et al., 2012; Kühn-Popp et al., 2013; Liu et al., 2004, 2009b, 2009a; Meinhardt et al., 2011, 2012) facilitated direct comparisons of neural responses associated with false belief understanding across adults and children by using ERP paradigms shared between studies. These studies fall into four groups, each using identical ERP paradigms for adults and children (see **Table 5**):

(1) Liu et al. (2004, 2009b) examined differences between belief and reality processing across adults and 4- to 6-year-old children. Participants were exposed to forty consistently structured cartoon animations displayed on a computer monitor. In these animations, a protagonist placed two animals into separate boxes and then exited the scene. One animal either switched boxes or returned to its original location during the protagonist's absence. After each scenario, participants were asked to make two judgments: a "reality judgment" concerning the current location of an animal and a "think judgment" about the protagonist's belief regarding the animal's location. The order of these judgments varied among participants, providing essential data for aligning ERP recordings that were critical for differentiating between reality and belief processing.

In adults, Liu et al. (2004, 2009b) observed that the late slow waveform divergence between reality and think conditions, occurring during the 700–900 ms post-stimulus interval, was primarily localized to the left frontal scalp region. This divergence was notably absent in the posterior and right hemisphere electrodes, indicating that reasoning about mental states versus reality in adults is predominantly associated with a late slow waveform in the left frontal electrodes. In contrast, children aged 4 to 6 years (Liu et al., 2009b) who passed the false belief judgments displayed significant differences between reality and thinking conditions across all frontal electrodes. These findings revealed that although these children also exhibited a left-lateralized late slow waveform, it was more diffuse and less distinctly lateralized than that observed in adults, highlighting developmental variations in neural processing strategies related to mental state reasoning.

(2) Meinhardt et al. (2011) investigated neural responses to true and false beliefs using non-verbal cartoon narratives based on the “Sally Anne Scenario” (Baron-Cohen et al., 1985). These narratives depicted protagonists (e.g., Betty and Nick) interacting with a target object (e.g., a teddy bear) and containers (e.g., a bag, a basket) across a series of seven pictures that branched into true and false belief conditions. The storyline commenced with Betty placing an object in a container and leaving the room, after which Nick would relocate the object. The narratives diverged when Betty returned: depending on whether she observed Nick’s actions, her search for the target object would be influenced, leading to either correct or incorrect search outcomes. These scenes aimed to trigger expectation violations, depicted in the final frames.

During the LPC (300-600 ms post-stimulus), adults exhibited a belief effect over central sites. In contrast, for children aged 6 to 8 years, differentiation between true and false belief reasoning was more pronounced at posterior sites. Moreover, the late waveforms extended over a longer duration for children (750-1450 ms post-stimulus) compared to adults (600-900 ms post-stimulus), indicating a more diffuse distribution of the belief effect encompassing the anterior regions at midline, superior, and inferior sites in children, whereas, in adults, it was more confined to the anterior regions at midline and superior sites.

(3) Bowman et al. (2012) and Liu et al. (2009a) conducted ERP studies designed to elicit diverse desires, diverse beliefs, and physical judgments. Each of the 48 identically structured trials per condition began by introducing participants to two protagonists with differing preferences for food or toys or two distinct storage locations, accompanied by corresponding images. One-third of the trials incorporated a memory check to maintain participant engagement and accuracy. Subsequently, participants addressed a target question relevant to the trial's theme—desires, beliefs, or physical locations—followed by a 2,000 ms display of an image of the pertinent item, during which ERP measurements were time-locked. Participants' responses involved selecting a character or location based on the revealed item. This consistent methodology across conditions was designed to ensure that any observed differences in neural responses were attributable to variations in mental-state processing rather than to perceptual or task-specific strategies. Trials were systematically blocked by condition and type to prevent repetitive patterns, thus optimizing the focus on cognitive processing distinctions.

In adults, Liu et al. (2009a) observed distinct late waveforms: between 800 and 850 ms post-stimulus, significant differences over midfrontal sites between belief and physical conditions, and between 600 and 800 ms post-stimulus, over right posterior sites. In children aged 7 to 8 years, Bowman et al. (2012) reported that the late waveform from 350 to 850 ms post-stimulus differentiated belief and physical conditions over midfrontal sites. Additionally, between 600 and 1,400 ms post-stimulus, differences were observed over right posterior sites, but only in trials where responses were correct.

(4) Kühn-Popp et al. (2013) and Meinhardt et al. (2012) investigated the neural correlates associated with reality, pretense, and false belief processing through ERP studies that adapted a behavioral task originally developed by Perner et al. (1994). Each cartoon scenario presented a protagonist interacting with an animal and its cage across three distinct conditions: the reality condition,

where the animal was visible inside the cage; the pretense condition, depicting the animal jumping out of the cage; and the false belief condition, where the animal jumped out unbeknownst to the protagonist. Participants observed these scenarios and subsequently responded to questions about the protagonist's behavior, such as "Why is he/she doing this?" The options for responses included "knowing" or "pretending," aimed at discerning the protagonist's perceived state of awareness. Critical ERP data were collected as participants responded, with control trials incorporated to mitigate anticipation effects. Responses were captured through a button press on a two-key response pad, with systematic variations in the gender of the protagonist and the type of animal featured.

In adults, Meinhardt et al. (2012) observed that during the late waveform period, 600 to 900 ms post-stimulus, the differences between false belief and reality conditions were predominantly localized over the anterior midline electrode FCz. Conversely, in children aged 6 to 8 years, Kühn-Popp et al. (2013) found distinctions in the late waveform from 600 to 920 ms post-stimulus over frontocentral sites between false belief and reality conditions. Additionally, a positive late waveform observable from 420 to 920 ms post-stimulus over parieto-occipital positions (PO7, PO8) further differentiated between false belief and reality conditions, highlighting developmental differences in the neural processing of false belief understanding. Furthermore, the findings from the studies by Kühn-Popp et al. (2013) and Meinhardt et al. (2012) on the neural underpinnings of pretense and false belief in both adults and children indicate that these forms of reasoning are supported by distinct neurocognitive mechanisms, suggesting that pretense and false belief are not functionally equivalent metarepresentational activities.

These studies revealed that both children and adults exhibit late waveforms in response to belief stimuli, yet they differ in scalp distribution and temporal characteristics. The observed distribution of late waveforms associated with false belief understanding mainly manifested in two brain regions; one component primarily situated in the frontal regions (e.g., Geangu et al., 2013; Liu et al., 2009b; Meinhardt et al., 2011, 2012; Zhang et al., 2013), while the other was predominantly found in the posterior area, often measured in parietal or parieto-occipital regions (e.g., Guan et al., 2020; Kühn-Popp et al., 2013; Meinhardt et al., 2011). Compared to adults, children's neural responses to belief information predominantly appear over parieto-occipital sites or show a more diffused frontal scalp distribution, often occurring later than in adults (Bowman et al., 2012; Kühn-Popp et al., 2013; Liu et al., 2009b; Meinhardt et al., 2011) (see **Figure 6**). These findings enhance our understanding of the developmental differences in brain activity related to belief reasoning between children and adults. They consistently demonstrate the presence of late frontal waveforms when reasoning about false beliefs, alongside developmental changes in the contribution of the parieto-occipital regions.

In early childhood, explicit false belief understanding emerges around ages 3½ to 5 years, as evidenced by traditional elicited-response false belief tasks (Wellman et al., 2001; Wimmer & Perner, 1983). Investigating the neural correlates within this critical age range is critical for elucidating the neurocognitive mechanisms underpinning false belief understanding. However, research with visual stimuli in young children presents challenges due to their limited attention spans, heightened motor activity, and increasing autonomy. To date, the study by Liu et al. (2009b) remains the only ERP investigation within this sensitive developmental window that compares neural responses to belief attributions and reality judgments between children aged 4 to 6 years and adults. In this study, children were categorized as "passers" or "failers" based on their performance in an independent behavioral false belief task. Notably, a late waveform over frontal regions was observed only in children who passed the false belief task, displaying a more diffuse scalp distribution compared to adults, while children who failed did not show consistent neural differentiation between belief and reality conditions.

These findings constitute the only existing evidence linking brain activity to the acquisition of false belief understanding in early childhood. Given the demonstrated behavioral competence of toddlers younger than 3 years in low-demands false belief tasks, it is crucial to further explore potential brain-behavior connections related to false belief competence in this younger age group. Study 1 of the present dissertation aims to extend the findings of Liu et al. (2009b) to toddlers, thereby enhancing our understanding of the development of false belief understanding from a neural perspective. To navigate the methodological challenges inherent in studying this age group, we developed a novel ERP paradigm, termed “Leo’s belief task.”

Leo’s belief task, an ERP paradigm, was designed to evaluate toddlers’ neural responses to a protagonist’s (Leo’s) true or false beliefs, paralleling the behavioral explicit false belief task from the Theory of Mind scale (Wellman & Bartsch, 1988; Wellman & Liu, 2004). In this task, toddlers were first shown the actual location of objects (e.g., a ball in a box) and then informed of Leo’s beliefs about the object’s location, either correctly or incorrectly (e.g., “Leo thinks that the ball is in the box/bucket”). Rather than predicting behavior, this ERP study focused on the neural response to the final spoken word in the sentence expressing Leo’s belief, serving as the event eliciting the ERP. This setup allows for direct neural comparisons between Leo’s true and false beliefs under consistent conditions where the positions of the box and bucket and Leo’s presence remain unchanged. We propose that Leo’s belief task is conducive to identifying the neural correlates of false belief processing in toddlers.

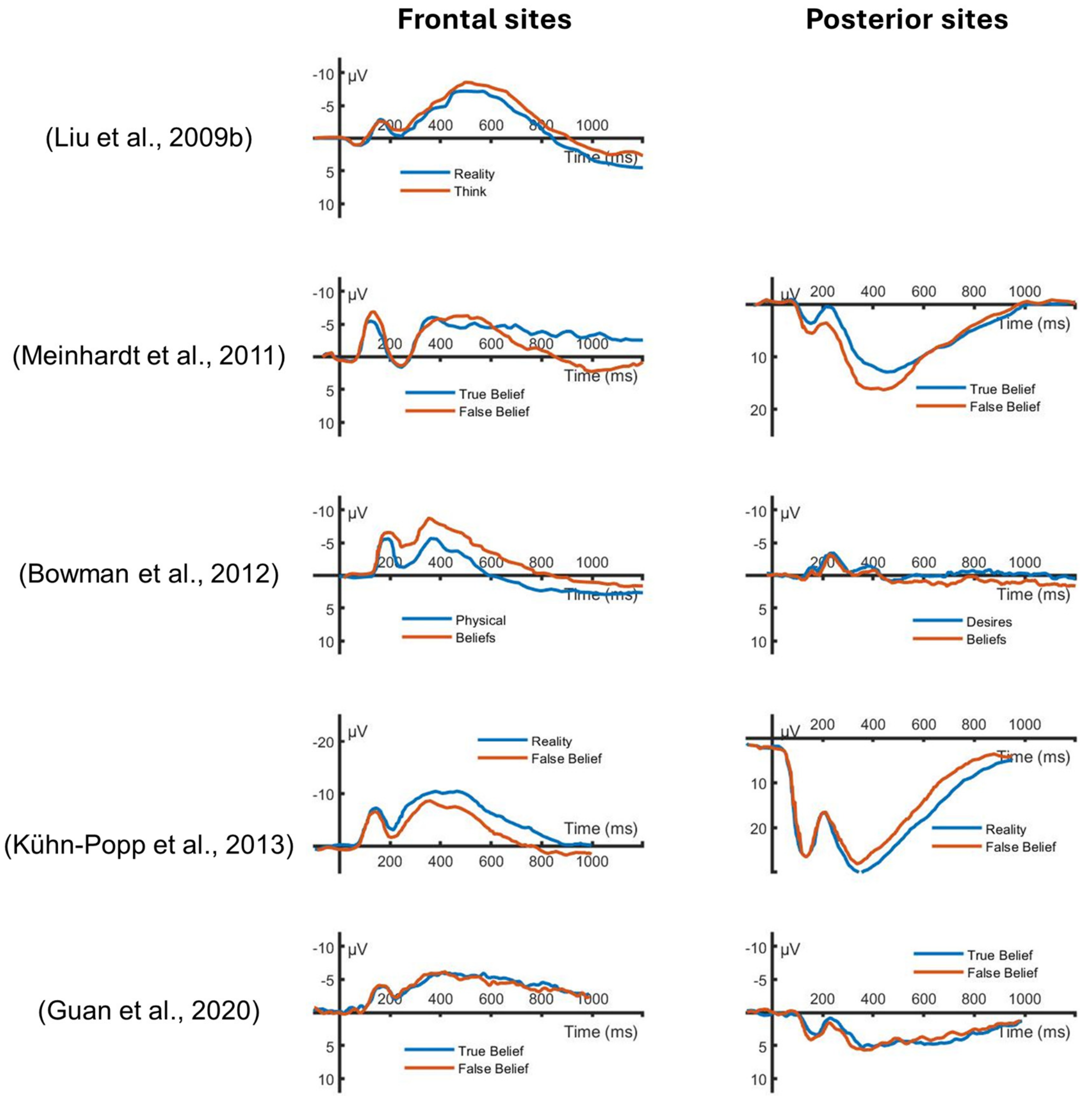


Figure 6. Comparison of late ERP waveforms across different Theory of Mind tasks in children aged 6-12 years.

Table 5. Comparison of ERP studies on false belief understanding between adults and children.

ID	Research	Sample size	Ages	Condition comparison	ERPs, Scalp distribution, Timing
1	(Liu et al., 2009b)	A: 24 C: 44	Adults: did not provide Children: 5 years 11 months	Belief/think condition vs. reality condition	Late slow waveforms: Adults: left frontal sites, 775-850ms Child passers: a more diffuse frontal scalp distribution, 1400-1500ms.
	(Liu et al., 2004)	A: 17	19-35 years	Belief/think condition vs. reality condition	Late slow waveforms: left frontal sites, 700-900ms.
2	(Meinhardt et al., 2011)	A: 21 C: 22	Adults: 24.33 years Children: 7.7 years	False belief vs. true belief	LPC, 300-600ms Adults: central sites; Children: parietal sites Late slow waveforms: Adults: midfrontal sites, 600-900ms. Children: broad anterior sites, 750-1450ms.
3	(Bowman et al., 2012)	C: 18	8.11 years	Belief reasoning vs. desire reasoning vs. physical control	Late slow waveforms: Belief/Desires vs. physical: mid-frontal slow sites, 200-250ms, 350-850ms Belief vs. Desire judgment: right posterior sites, 600-800ms, 850-1400ms
	(Liu et al., 2009a)	A: 15	22 years	Belief reasoning vs. desire reasoning vs. physical control	Late slow waveforms: Belief/desire vs. physical: midfrontal sites, 800-850ms. Belief vs. physical: right posterior sites, 700-800ms. Belief vs. physical, belief vs. desire: right posterior sites, 600-700ms.
4	(Kühn-Popp et al., 2013)	C: 21	7.34 years	False-belief reasoning vs. reality	Late slow waveforms: frontocentral sites, 600-920ms; parieto-occipital sites, 420-920ms.
	(Meinhardt et al., 2012)	A: 24	23.5 years	False-belief reasoning vs. reality	Late slow waveforms: midfrontal sites, 600-900ms

Note: A = adults; C = children. LPC = late positive complex.

1.5. The potential relationship between task-independent neural systems and Theory of Mind

A growing body of evidence suggests that children's explicit false belief understanding may develop gradually over time rather than through abrupt, fundamental conceptual shifts during early childhood (Richardson & Saxe, 2020a). If Theory of Mind is indeed a progressive developmental phenomenon, with conceptual improvements throughout early childhood, the question of how functional brain specializations support these sensitive developments remains open.

As overviewed earlier, task-dependent brain activity measured by MRI and neural responses assessed using ERP techniques reveal specialized neural systems and some typical waveforms involved in Theory of Mind. In addition to these task-dependent processes, the brain also includes regions that are typically more active during rest compared to active task performance. These regions include the default mode network (Mars et al., 2012a), identified through MRI, as well as resting-state brain activity indices (e.g., brain asymmetry and coherence) measured using baseline EEG (Sabbagh, 2013). This section reviews the literature on the potential relationship between these task-independent neural systems and Theory of Mind.

1.5.1. *Whether functional specializations are a cause or a consequence of Theory of Mind development*

Existing neuroimaging studies found that as specific neural systems—including the MPFC and the right TPJ—become increasingly engaged in mental state reasoning rather than in general social information (Schurz et al., 2014, 2017, 2021), children demonstrate improved performance on behavioral Theory of Mind tasks. This reflects a growing "selectivity" of these regions for mental state reasoning. Even in the absence of external stimuli, these regions function as an integrated network known as the default mode network, which shows a remarkable overlap with the brain regions associated with Theory of

Mind (Mars et al., 2012a; Schilbach et al., 2008; see **Figure 7**).

The default mode network is frequently identified in studies utilizing a "resting-state" paradigm. In these paradigms, participants' brain activity is recorded while they are not engaged in any specific task, allowing them to think freely without directed focus (Buckner et al., 2008; Raichle et al., 2001). Under these conditions, a distinct set of brain regions—including the MPFC and the right TPJ—consistently show activity (M. D. Fox et al., 2005; Raichle et al., 2001; Sridharan et al., 2008). Recent research has further demonstrated a remarkable overlap between the brain regions typically involved in Theory of Mind processes and the default mode network (for a review, see Mars et al., 2012a), a finding that warrants further attention. For example, studies found that lateral nodes are linked to understanding others' mental states (Koster-Hale et al., 2017; Sridharan et al., 2008). Specifically, the right TPJ is selectively involved in representing others' beliefs, desires, and intentions (e.g., Aichhorn et al., 2009; Perner et al., 2006; Saxe et al., 2009). Functional neuroimaging research in neurotypical adults highlights the selective recruitment of the right TPJ in attributing mental states (Saxe & Powell, 2006; Saxe & Wexler, 2005) and in discerning false beliefs from closely related scenarios (Perner et al., 2006; Saxe & Kanwisher, 2003). Nonverbal cartoon stimuli also engage right TPJ activity specifically when interpreted in terms of a character's false beliefs (Saxe et al., 2006). Further fMRI studies with school-aged children (ages 5 to 11) suggest that the right TPJ increasingly distinguishes mental states from physical facts as children age (Gweon et al., 2012; Saxe et al., 2009). Thus, Given that the default mode network overlaps with task-dependent brain activity during Theory of Mind tasks, it suggests that humans are predisposed to social cognition as a default mode of cognitive processing, supported by the intrinsic activity of the "default system." (Schilbach et al., 2008).

Most of the existing literature on the default mode network focuses on adult subjects, and a comprehensive understanding of default mode network development during early childhood—when the brain undergoes tremendous development—is still lacking. In default mode network studies conducted in infants, Gao et al. (2009, 2013) found that response time courses in MPFC and PC regions in were observed in 1-year-old infants and correlated responses in MPFC and TPJ in 1- and 2-year-olds. A recent study by Xiao et al. (2016) investigated the development of core regions and subsystems of the default mode network in children aged 3 to 5 years. Significant changes were observed in the posterior subsystem, while the frontal subsystem remained unchanged, suggesting distinct developmental trajectories. They also found stronger right hemispheric lateralization at age 3, which gradually shifted towards more bilateral development by age 5, suggesting that hemispheric dominance changes with age. These results provide preliminary evidence for the development of default mode network subsystems in early life. However, whether this development is closely related to the maturation of Theory of Mind in childhood remains unclear, and whether the functional specializations are a cause or a consequence of Theory of Mind development is still an open question.

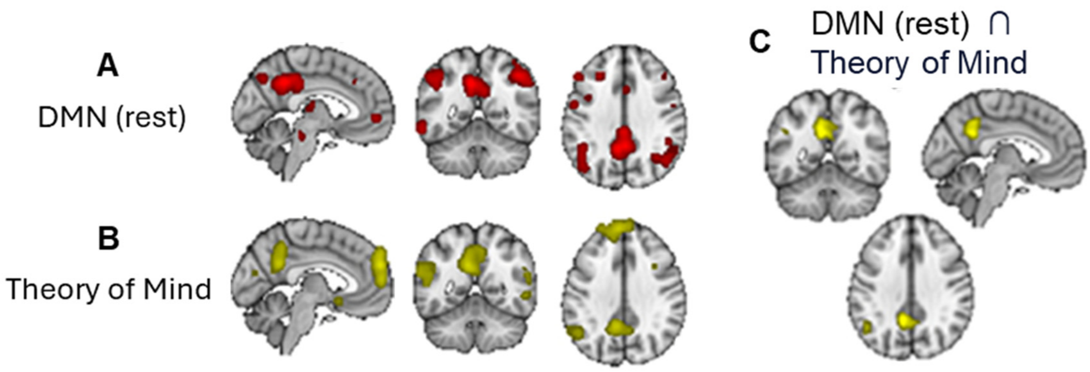


Figure 7. Overlap between the default mode network (DMN) and areas activated by Theory of Mind paradigms (adapted from Mars et al., 2012a). Activation likelihood maps of activity during (A) passive resting-state conditions and (B) Theory of Mind. (C) Conjunction maps of resting-state and Theory of Mind showed significant overlap in regions such as the angular gyrus, posterior cingulate cortex, and precuneus.

1.5.2. Resting-state electroencephalogram measures of neurocognitive development

Encephalographic recordings have been employed for nearly a century to provide insights into brain activity and its involvement in cognitive functions (see Millett, 2001, for a review). This technique is widely used in cognitive neuroscience due to its cost-effectiveness and non-invasive nature, which imposes minimal physical strain on participants. Previous research has utilized resting-state EEG power in specific frequency bands during baseline conditions to evaluate the functional development of various brain regions in children (Bell, 2001, 2002; Bell & Fox, 1992, 1997). Research has shown that from late infancy through early childhood, alpha waves exhibit a clear developmental increase in resting EEG rhythm across multiple scalp regions, with a peak frequency of 6–9 Hz in infants and young children (Marshall et al., 2002). When the brain is at rest, alpha frequency brain waves become more synchronized and increase in amplitude across the scalp, as measured by resting-state EEG. In contrast, when the brain is actively engaged in cognitive tasks, alpha wave activity becomes less synchronized and decreases in strength (Klimesch, 1999). As a result, alpha wave power serves as a reliable indicator of brain activity, with reduced alpha power typically reflecting greater cognitive engagement (Gevins, 1998).

Resting-state EEG alpha power is frequently used to assess individual differences in tonic cortical activation (MacLean et al., 2012; Pitchford & Arnell, 2019). Researchers particularly focused on how regional variations in tonic cortical activation may serve as a trait-like characteristic that predisposes individuals to cognitive development across various domains, including psychopathology, motivation, and personality (N. A. Fox et al., 1995; Licata et al., 2015; Müller et al., 2015, 2018; Paulus, Kühn-Popp, et al., 2013; Van Der Vinne et al., 2017). Of particular interest in this regard has been coherence and hemispheric asymmetry in frontal alpha power. Resting-state EEG alpha coherence is used to assess the degree of integration processes within the brain (Aykan et al., 2021). Resting-state EEG alpha asymmetry measures the difference in alpha power between the two brain hemispheres (Müller et al., 2015). Alpha waves are thought to relate inversely to brain activity (Gevins, 1998; Klimesch, 1999), implying that higher alpha power in one hemisphere suggests increased activity in the opposite hemisphere (Reznik & Allen, 2018). Researchers measure resting-state EEG alpha asymmetry by subtracting left alpha power from right alpha power, where negative values indicate dominant right hemisphere activity, and positive values indicate dominance in the left hemisphere (Allen et al., 2004; Müller et al., 2015, 2018; Sabbagh & Flynn, 2006). The regional decreases in resting/baseline alpha power (e.g., Marshall et al., 2002; Thatcher, 1992, 1994; Thatcher et al., 1987) are thought to reflect the more mature functional organization of the underlying neurocognitive systems (Sabbagh et al., 2009; Thatcher, 1994). For example, consistent findings were observed in adult populations. Sabbagh and Flynn (2006) examined whether individual differences in resting frontal EEG alpha asymmetry predicted performance in mental-state decoding, a key component of Theory of Mind. Their findings indicated that individuals with right-lateralized frontal activation exhibited stronger mental-state decoding skills, and the degree of right mid-frontal activation was a significant predictor of Theory of Mind performance. This finding suggests a potential trait-like aspect of resting-state EEG alpha asymmetry in Theory of Mind abilities. Additional evidence supporting the right hemispheric lateralization of brain activity associated with false belief understanding comes from a resting-state EEG source-localized analysis (Sabbagh et al., 2009). Sabbagh et al. (2009) identified individual differences in alpha oscillations linked to Theory of Mind performance in 4-year-old children. Specifically, regions in the dorsal MPFC and several right hemisphere areas—including the TPJ, precentral gyrus, cuneus, and inferior temporal cortex—were associated with the representational Theory of Mind.

These associations are held after controlling for children's executive functioning and language skills. These findings were interpreted as evidence that the maturation of these regions may underpin the development of false belief understanding and related Theory of Mind abilities. A further longitudinal study based on Sabbagh's work examined whether source-localized resting-state EEG activity in the dorsal MPFC or the right TPJ at age 4 could predict ToM-specific fMRI responses 3.5 years later. The findings revealed that preschoolers' resting-state EEG activity in the dorsal MPFC predicted subsequent ToM-specific fMRI responses in the same region (Bowman et al., 2019). These findings have implications for characterizing the conceptual Theory of Mind development and its underlying neural supports.

Findings on cognitive tasks showed that resting-state EEG alpha asymmetry can predict task performance in a manner consistent with lesion and neuroimaging studies (Hoptman & Davidson, 1998). Approximately 60% of the variance in alpha asymmetry indices is thought to represent a latent trait with near-perfect temporal stability, which is linked to various psychological constructs (N. A. Fox et al., 1995; Hagemann et al., 2002; Stewart et al., 2011). Research indicates that resting-state EEG frontal alpha asymmetry is a stable characteristic over time across various age groups in children (N. A. Fox et al., 1992; Jones et al., 1997; Müller et al., 2015). For instance, studies reported its stability over a 24–30 month span in early childhood aged 3-6 (Jones et al., 1997), over a 6–36 month span in children aged 3–5 (Vuga et al., 2008), and over a 4-year span in children aged 4–8 (Kim & Bell, 2006). Furthermore, similar findings were observed over extended periods of up to 69 months across infants and preschoolers. For example, Müller et al. (2015) demonstrated high individual stability of frontal alpha asymmetry spanning from 14 to 83 months of age. Collectively, these findings suggest a consistent pattern of stability in frontal alpha asymmetry across different developmental age ranges. Although these neuroscientific studies indicate that resting-state EEG alpha asymmetry may be a potential stable neural marker during early childhood, the existing evidence linking resting-state EEG alpha asymmetry with Theory of Mind competencies in children remains limited.

1.5.3. Children's brain asymmetric activity associated with false belief understanding development

Although we provided an overview of the potential relationship between task-independent neural systems and Theory of Mind, research in this field during early childhood remains limited. Based on the task-dependent neural correlates of false belief understanding in children presented in Section 1.3.3, findings suggest an early developmental origin of a mentalizing network and potential right hemisphere brain activity during Theory of Mind tasks. For example, Grosse Wiesmann et al. (2017b) utilized structural MRI to show that false belief competence in 3- and 4-year-olds correlates with age-related increases in local white matter in regions such as the right ventral MPFC, right TPJ, and right PC. Hyde et al. (2018) tested the specific hypothesis that the right TPJ is preferentially active when thinking about others' mental states, even in 7-month-old infants. Moreover, evidence that the default mode network overlaps with brain regions associated with Theory of Mind (Mars et al., 2012a), along with a recent study showing stronger right hemispheric lateralization at age 3 in children aged 3 to 5 years (Xiao et al., 2016), and findings from 1-year-old infants there showing resting-state activity in the MPFC and TPJ regions (Gao et al., 2009, 2013), all suggest that exploring candidate neural markers under resting-state conditions would be useful for understanding the drivers of individual differences in Theory of Mind. This could help determine whether functional specialization is a cause or a consequence of representational Theory of Mind development (Sabbagh et al., 2009).

Another line of evidence for the relevance of task-independent brain asymmetry to Theory of Mind came from studies involving individuals with autism spectrum disorder, who typically show deficits in Theory of Mind abilities

(Baron-Cohen, 2000, 2001; Baron-Cohen et al., 1985; Happé & Frith, 1995; Leekam & Perner, 1991). For example, Stroganova et al. (2007) reported atypical broadband resting-state EEG asymmetry in children with autism aged 3 to 8 years, revealing reduced right temporal cortex activity in generating EEG rhythms. Thus, exploring candidate neural markers under resting-state conditions could be crucial for understanding the factors that promote or hinder the development of ToM-related brain regions, as well as for designing and assessing the effectiveness of clinical interventions aimed at improving socio-cognitive abilities (Richardson & Saxe, 2020a).

Given that false belief understanding is often regarded as a “litmus test” for evaluating Theory of Mind abilities in children (e.g., Wellman & Woolley, 1990), investigating how early neural characteristics, such as asymmetric activity observable in the resting state, relate to the developmental trajectory of false belief understanding in typically developing children is important. Longitudinal research is necessary to investigate the psychological and biological factors underlying developmental processes in early childhood (Bergman et al., 1989). To address this issue, researchers should collect longitudinal data by measuring task-independent brain activity in infancy or toddlerhood and subsequently assessing behavioral false belief understanding at later developmental stages. The individual differences in early brain activity may reflect stable neural traits over time that relate to false belief understanding, and early measurements of these neural markers should predict subsequent false belief abilities (Richardson & Saxe, 2020a).

Briefly summarized, Theory of Mind develops gradually, with cumulative evidence indicating that behavioral competence in false belief tasks emerges at earlier ages than previously assumed. The question of how the brain supports Theory of Mind during early development has not been fully resolved (Sabbagh et al., 2009). On the one hand, the neural specializations observed in older children and adults might reflect the consequences of developmental progression rather than being present from the outset (Elbert et al., 2001; Karmiloff-Smith, 1997; Sabbagh & Flynn, 2006). On the other hand, certain components of the Theory of Mind neural system may exhibit early specialization and longitudinal consistency, with individual differences in early specialization predicting later individual differences in the same regions (Bowman et al., 2019).

One of the aims of this dissertation is to explore the potential longitudinal relationship between infants' and toddlers' task-independent baseline EEG measures and preschoolers' behavioral competence in Theory of Mind tasks, thereby preliminarily filling a gap in our understanding of the early neural mechanisms underlying false belief understanding. By examining resting-state EEG alpha asymmetry, we seek to explore whether individual differences in brain asymmetries—particularly in the frontal and parietal regions—are evident before the emergence of false belief abilities or whether they develop as a correlate of mastering explicit false belief understanding. This investigation offers new insights into the neurodevelopmental foundations of conceptual Theory of Mind. Moreover, this approach allows us to explore the relationship between task-independent resting-state EEG and task-dependent behavioral competence, an area that is currently debated and remains unclear (e.g., Deco et al., 2011; Papo, 2013).

1.6. The present study

1.6.1. Research questions

As outlined in the previous sections, around age 4 is considered a sensitive period for the development of children's traditional explicit false belief understanding. However, recent research has begun to challenge the traditional view that this competence emerges only at this age. Some studies (Grosso et al., 2019; Setoh et al., 2016; Sodian et al., 2024) demonstrated that 2.5-year-old toddlers exhibit explicit false belief competence when tested with modified versions of traditional false belief tasks that reduce processing demands related to inhibitory control and response generation. Furthermore, studies employing anticipatory looking, violation-of-expectation, and prompted action paradigms suggest that even one-year-old infants can make action predictions based on others' beliefs (Buttelmann et al., 2009; Onishi & Baillargeon, 2005; Southgate et al., 2007). If false belief understanding exists on the behavioral level, it should also be reflected on the neural level. Therefore, it is plausible that neural activity supporting false belief understanding emerges as early as in toddlers or even in infants.

First, we propose that if children exhibit false belief understanding behaviorally before age 3, this may indicate an earlier onset of the cognitive mechanisms underlying false belief understanding, which could be reflected in their neural responses. The presence of distinct neural patterns between false belief and true belief conditions in younger children who pass the behavioral false belief task could indicate that these children developed a neural foundation for false belief understanding similar to that of slightly older children (ages 4–6) (Liu et al., 2009b). In essence, children who demonstrate false belief understanding at a younger age may also show a corresponding neural distinction, reflecting the early development of cognitive processes involved in false belief understanding.

Second, given the existing evidence supporting the early developmental origins of a specialized mentalizing network (Richardson & Saxe, 2020b; Schurz et al., 2021), an open question remains: How does the specialized neural activity associated with false belief understanding develop? This network may emerge as early as age 3, exhibiting highly correlated responses to mental state stimuli (e.g., Grosse Wiesmann et al., 2020; Richardson et al., 2018; Richardson & Saxe, 2020) and even in the absence of external stimuli (Mars et al., 2012a; Xiao et al., 2016). More specifically, studies indicated potential right-hemispheric lateralization of both task-dependent and task-independent brain activity associated with false belief understanding in infants and toddlers (Hyde et al., 2018; Sabbagh et al., 2009; Xiao et al., 2016). Are brain–behavior connections primarily driven by cognitive development, such that as children acquire the ability to solve false belief tasks, their neural responses to these scenarios adapt accordingly? Alternatively, is there longitudinal evidence that early neural activity patterns may predict later individual differences in false belief competence?

In this context, the present dissertation investigates the following overarching research questions:

1. What neural-behavioral connections underpin false belief understanding in toddlers younger than three years?

2. How does resting-state brain asymmetric activity associated with false belief understanding emerge? Is it a consequence of developing behavioral false belief competence, or does evidence of this system exist at the neural level prior to observable behavioral false belief competence?

1.6.2. Research design

The present doctoral dissertation addressed its research questions through three studies, as detailed in **Figure 8**, which outlines the research framework:

Study 1: This study explored the brain-behavior connections associated with false belief understanding in toddlers aged 33 to 36 months. Previous research by Setoh et al. (2016) challenged the traditional view that a representational Theory of Mind emerges only around age four, suggesting that toddlers approximately 2.5 years old can comprehend traditional explicit false belief tasks when processing demands are sufficiently reduced. To further investigate the neural correlates of false belief understanding, 33- to 36-month-old toddlers participated in the low-demands false belief task by Setoh et al. (2016) and a novel ERP task (Leo's belief task) designed to assess neural responses to belief understanding. Concurrently, co-developing cognitive abilities, including IQ and general language abilities, were measured. This study primarily addresses the first research question.

Study 2: Employing a longitudinal design, Study 2 examined the relationships between early resting-state EEG alpha asymmetries and false belief understanding during the critical developmental window between 3 and 4 years. Resting-state EEG measurements were taken at 34 months (Time 1), with subsequent assessments of explicit false belief tasks (content and location false belief tasks) and executive functions at 52 months (Time 2). Implicit false belief tasks, as outlined by Grosse Wiesmann et al. (2017a) and Kaltefleiter et al. (2022), along with general language abilities, were assessed at both time points. This study explores how changes and continuities in neural mechanisms contribute to the development of false belief understanding during early childhood.

Study 3: Building on the preliminary findings from Study 2 regarding explicit false belief understanding, this study incorporates an independent longitudinal dataset to replicate and substantiate the relationship between resting-state EEG alpha asymmetry and explicit false belief understanding across an extended age interval. Resting-state EEG was recorded at 14 months (Time 1), with follow-up assessments of explicit false belief tasks, general language abilities, and executive function conducted at 51 months (Time 2). Studies 2 and 3 collectively address the second research question.

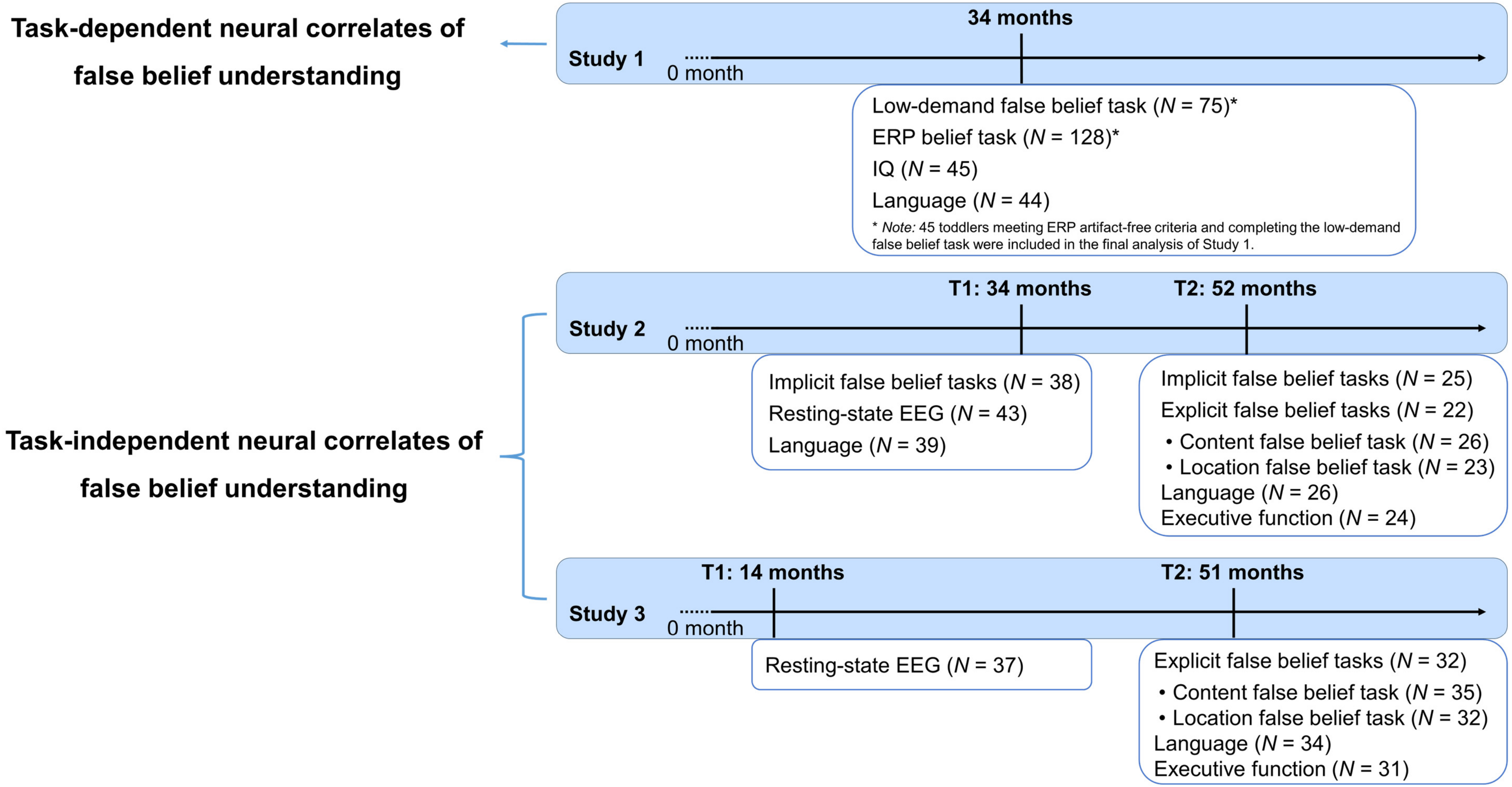


Figure 8. Overview of the research design and studies.

Chapter 2: Neural Correlates of False Belief Understanding in 33- to 36-Months-Old Toddlers (Study 1)

Highlights

1. Brain-behavior connections associated with false belief understanding emerged in toddlers under the age of three years.
2. Toddlers who passed a low-demands behavioral false belief task showed a late positive waveform over the occipital electrode sites, which differentiated between the false belief and true belief conditions.
3. The association between one of two late waveforms and behavioral false belief competence raises the possibility that a sensitive neural system of false belief understanding may emerge early in development.

Abstract

Very little research has addressed the neural correlates of false belief understanding in young children. Following up on a study by Liu et al. (2009b) in 4-to 6-year-olds, the present study grouped toddlers (*N* = 45, 33-36 months old) into passers and failers according to their performance on a behavioral false belief task with reduced processing demands (see Setoh et al., 2016). Their event-related potential (ERP) responses to the false belief and true belief conditions were examined in a novel ERP paradigm. This study found that a late positive waveform over the occipital electrode sites distinguished between the false belief and true belief conditions only in toddlers who passed the low-demands behavioral false belief task. In contrast, a late negative waveform over the frontocentral electrode sites consistently distinguished between the false belief and true belief conditions regardless of low-demands behavioral false belief task performance. These findings raise the possibility that a sensitive neural system supporting false belief understanding may emerge early in development. Specifically, the late positive waveform observed over the occipital electrode sites appears to be a potential neural marker for false belief understanding in toddlers.

Keywords

False belief understanding, event-related potential, brain-behavior connections, mentalizing, toddlers.

Dataset source and Shuting Li's contributions to the study

Li, S., Meinhardt, J., Sodian, B. (under review). Neural Correlates of False Belief Understanding in 33- to 36-month-old Toddlers. Study 1 is part of a larger project on early Theory of Mind development, *"The Role of Language in Early Theory of Mind Development"* (DFG SO 213/33-1,2). Jörg Meinhardt and Beate Sodian designed a novel ERP paradigm for toddlers and supervised data collection. Shuting Li analyzed the data and integrated them with behavioral measures of false belief understanding. She took the intellectual lead in conceptualizing the paper, conducting data analysis, visualizing the results, and framing and writing the manuscript.

2.1. Introduction

Theory of Mind, which refers to the comprehension of mental states underlying behavior, plays a pivotal role in cognitive development. A crucial aspect of the Theory of Mind is false belief understanding, denoting the ability to discern between reality and an agent's false mental representation of the world (Tomasello, 2018; Wimmer & Perner, 1983). It involves grasping the causal link between people's access to information and their mental representations, as well as the causal influence of beliefs on actions (Wellman et al., 2001). Typically-developing children begin to succeed in attributing false beliefs to themselves and others between approximately 3.5 and 5 years of age (Perner, 1991; Wimmer & Perner, 1983). With the acquisition of false belief understanding, children realize that beliefs are subjective mental representations rather than reflections of reality, and that people act on their beliefs (Scott & Baillargeon, 2017). In summary, false belief understanding is a milestone of socio-cognitive development in general, and of Theory of Mind in particular, since it is critical to understanding, explaining, and predicting behavior (Wellman, 2010; Wellman et al., 2001).

Over the last two decades, significant progress has been made in identifying the neural correlates of false belief understanding. Research using functional magnetic resonance imaging (fMRI) has shown activation of two main brain regions during false belief tasks, including the MPFC and TPJ (Döhnel et al., 2012; Le Petit et al., 2022; S. Li et al., 2023; Saxe, 2009; Saxe et al., 2004a; Schuwerk, Grosso, et al., 2021; Sommer et al., 2007). The event-related potential (ERP) technique's high temporal resolution has enabled examining the time course of neural responses (Sabbagh, 2013). Late waveforms, which can be negative or positive, emerge around 600 milliseconds post-stimulus and continue to the end of one second of the recording epoch (Sabbagh, 2013). Variations in the topography, amplitude, and temporal duration of these waveforms are closely associated with the cognitive subprocesses required by specific task demands. Previous ERP studies of mentalizing found a late waveform effect associated with false belief processing (e.g., Liu et al., 2004; Sabbagh & Taylor, 2000). This effect has been replicated with many different kinds of stimuli (e.g., Geangu et al., 2013; Kühn-Popp et al., 2013; Liu et al., 2009b, 2009a; Meinhardt et al., 2011, 2012). In Theory of Mind tasks, the late waveform typically manifests extensively across the frontal and parietal electrode sites, a pattern frequently observed when participants engage with complex stimuli that require the attribution of mental states to others.

Late waveforms were often considered a neural marker of the mental operations involved in false belief understanding. The late waveforms involved in false belief understanding were often characterized as "late slow waveforms" (e.g., Bowman et al., 2012; Geangu et al., 2013; Guan et al., 2020; Meinhardt et al., 2012). Slow waveforms are tonic with less distinct peaks, persisting from several hundred milliseconds to several seconds (Rösler et al., 1997). In ERP studies on the neural correlates of false belief understanding with adults and children, the term late slow waveform has been used widely and not quite consistently. One aspect of the late slow waveform showed differences in amplitude between mental state attribution processes and non-mental state attribution processes (e.g., Bowman et al., 2012; Guan et al., 2020; Liu et al., 2009a, 2009a; Sabbagh & Taylor, 2000). For example, false belief reasoning, as compared to false photo reasoning, exhibited differences in the amplitude of a left-frontal late slow waveform, observed between 600 and 840 milliseconds post-stimulus (Sabbagh & Taylor, 2000). Belief reasoning, encompassing both false belief and true belief, versus reality understanding elicits a more negative late slow waveform over left-frontal sites from around 700 to 900 ms in adults (Liu et al., 2004, Liu et al., 2009b), whereas in children it appears from 1,400 to

1,500 ms (Liu et al., 2009b). Another aspect of variations in the late slow waveform was observed across different mental state processes. For instance, the late slow waveforms for false belief reasoning differ from those observed in true belief reasoning (e.g., Chen et al., 2012; Geangu et al., 2013; Meinhardt et al., 2011; Zhang et al., 2013), desire reasoning (Bowman et al., 2012; Liu et al., 2009a), and pretense (Kühn-Popp et al., 2013; Meinhardt et al., 2012). Compared to true belief reasoning, false belief reasoning elicited a late slow waveform with greater amplitude over midfrontal sites between 600 and 1000 ms in adults (Chen et al., 2012; Geangu et al., 2013; Meinhardt et al., 2011; Zhang et al., 2013). In children, false belief reasoning similarly elicited a late slow waveform with greater amplitude than true belief reasoning; however, this effect was observed at broadly distributed anterior sites and occurred later, between 750 and 1450 ms (Meinhardt et al., 2011). In previous literature, late slow waveforms were proposed to reflect processes associated with decoupling mental states from reality (e.g., Liu et al., 2004) or the decoupling mechanism involved in meta-representation (Kühn-Popp et al., 2013; Meinhardt et al., 2012). Additionally, they were linked to working memory processes (Bailey et al., 2016; Barriga-Paulino et al., 2014), suggesting they may reflect extended working memory processing and allocation of attention resources during belief reasoning.

Developmental investigations comparing neural activation between 6- to 12-year-old children and adults consistently show that children demonstrate a more diffuse distribution of late slow waveforms compared to adults (e.g., Guan et al., 2020; Kühn-Popp et al., 2013; Liu et al., 2009b; Meinhardt et al., 2011). Seven ERP studies (Bowman et al., 2012; Kühn-Popp et al., 2013; Liu et al., 2004, 2009b, 2009a; Meinhardt et al., 2011, 2012) facilitated direct comparisons of neural responses associated with false belief understanding across adults and children by using ERP paradigms shared between studies.

These studies fall into four groups, each using identical ERP paradigms for adults and children (see supplementary materials S1): (1) Meinhardt et al. (2011) focused on neural responses to false and true belief; (2) Liu et al. (2004, 2009b) explored belief and reality; (3) Kühn-Popp et al. (2013) and Meinhardt et al. (2012) examined false belief understanding, reality, and pretense; (4) Bowman et al. (2012) and Liu et al. (2009a) analyzed belief, desire, and physical control conditions. These studies revealed that both children and adults exhibit late slow waveforms to belief stimuli yet differ in scalp distribution and temporal characteristics. In adults, late slow waveforms related to belief processing typically manifest over frontal sites (Liu et al., 2004, 2009a; Meinhardt et al., 2011, 2012) or right posterior sites (Liu et al., 2009a) during the 600-900 ms interval. Conversely, children's neural responses to belief information predominantly present a more diffused frontal scalp distribution or appear over parieto-occipital sites, often occurring later than in adults (Bowman et al., 2012; Kühn-Popp et al., 2013; Liu et al., 2009b; Meinhardt et al., 2011). These findings enhance our understanding of the developmental differences in brain activity related to belief reasoning between children and adults. They consistently demonstrate the presence of frontal late slow waveforms in reasoning about false beliefs. Additionally, there are developmental changes in the contribution of the parieto-occipital regions. The present study, which assessed the neural responses to false belief, hypothesizes that in toddlers, these responses manifest as more diffused and less localized late waveforms, potentially encompassing broader frontal and parieto-occipital areas and occurring at a later time window.

In behavioral child development, false belief understanding has been shown to emerge with steep developmental progress between 3.5 and 5 years of age in traditional behavioral false belief tasks (Wellman et al., 2001; Wimmer & Perner, 1983). Investigating the neural correlates of false belief understanding within this sensitive age range is essential for understanding the neurocognition of false belief understanding. However, at this age, limited attention spans, increased motor activity, and growing autonomy lead to difficulties in maintaining engagement and adherence to ERP experiment protocols. Only one ERP study has examined children's belief reasoning in this sensitive age range. Liu et al. (2009b) compared the neural responses of children aged 4 to 6 years and adults in belief attributions to story characters and reality judgments. Children were grouped into passers and failers based on an independent assessment of their behavioral false belief competence. A late slow waveform was observed over frontal sites only for false belief passing children, with a more diffuse scalp distribution than that observed in adults. Children who failed the behavioral false belief task showed no systematic differentiation between belief and reality conditions on the neural level. To date, these findings provide the only evidence for brain-behavior connections in acquiring false belief understanding in early childhood.

Recent behavioral research shows that toddlers younger than 3 years can pass explicit verbal false belief tasks with reduced processing demands, indicating an early false belief understanding and conceptual continuity (Grosso et al., 2019; Scott et al., 2020; Setoh et al., 2016), rather than conceptual change over the preschool years. However, no prior studies have investigated the neural correlates of explicit verbal false belief understanding in this age group. Understanding these neural bases is important for tracing the developmental origins of false belief understanding on the neural level. The present study builds upon Liu et al. (2009b) by extending their findings to toddlers, thereby broadening our understanding of the developmental origins of false belief understanding from a neural perspective. To address methodological challenges in this age group, we developed a novel ERP paradigm (i.e., Leo's belief task, see below) and we employed the low-demands behavioral false belief task adapted from Setoh et al. (2016).

The low-demands behavioral false belief task helps toddlers overcome response-generation and inhibitory-control difficulties. Response generation involves toddlers' interpreting, retaining, and responding to a standard question that predicts the protagonist's behavior. Inhibitory control requires them to suppress their knowledge of reality to accurately reflect the protagonist's false belief. In Setoh et al.'s (2016) study, toddlers were told a story about protagonist Emma, who found an apple in one of two containers, moved it to the other container, and went outside. In her absence, her brother found the apple and took it away. Emma then returned to look for her apple. In the test trial, toddlers were shown pictures of the two containers and were asked the test question, "Where will Emma look for her apple?" The test object was removed from the scene to reduce demands for inhibitory control, obviating the requirement for toddlers to overcome reality bias (Robinson & Mitchell, 1995). Importantly, toddlers were also given two practice trials to reduce demands on response generation. In one practice trial, toddlers saw an apple and a banana and were asked, "Where is Emma's apple?" in the other practical trial, they saw a ball and a frisbee and were asked, "Where is Emma's ball?". Under these conditions, 30- to 33-month-old toddlers performed above chance in the test trial; 25 of 32 (78%) toddlers pointed to the container that Emma mistook to hold her apple (Setoh et al., 2016, Exp. 1). Grosso et al. (2019) replicated this finding in 33-month-old toddlers. Importantly, Scott et al. (2020) found false belief competence in toddlers also in a false belief about identity task with low demands. Moreover, a recent longitudinal study found a longitudinal correlation between false belief understanding assessed with the low-demands task adapted from Setoh et al. (2016) at 33 months and false belief understanding in a standard false belief assessment at 52 months (Sodian et al., 2024). These findings support the view that there is conceptual continuity in false belief understanding from infancy to preschool age and that toddlers' failure in traditional false belief tasks may be due to higher processing demands (Baillargeon et al., 2010, 2016). If there is conceptual continuity in false belief understanding on the behavioral level, this should be reflected on the neural level. Specifically, toddlers who pass the behavioral false belief task before the age of 36 months should also show a distinction between the false belief and true belief conditions on the neural level, similar to the distinction between passers and failers in 4- to 6-year-olds in Liu et al.'s (2009b) study.

The present study was designed to systematically investigate the neural correlates of false belief versus true belief processing in toddlers under 36 months of age. We developed an ERP paradigm named "Leo's belief task" to assess toddlers' brain responses as they process information about a protagonist's (i.e., Leo's) true or false beliefs. This paradigm (**Figure 9**) is akin to the behavioral "explicit false belief task" (Wellman & Bartsch, 1988; Wellman & Liu, 2004), which is part of the Theory of Mind scale. In Leo's belief task, toddlers were explicitly informed about reality (e.g., toddlers saw that the ball was in the box while the bucket was empty). Subsequently, toddlers were informed by verbal communication about Leo's true or false beliefs (e.g., "Leo thinks that the ball is in the box/bucket"). This information was followed by a behavioral test question: "Where will Leo look for (the target object)?". The explicit false belief task has been found to be equally challenging as other standard false belief tasks that require children to infer the protagonist's false belief from their access to information (Wellman & Bartsch, 1988). In the present ERP paradigm, however, we did not assess the neural response to a prediction of behavior, but we chose the final spoken word in the sentence expressing Leo's belief as an ERP eliciting event. This spontaneous response assessment enables a direct comparison of neural responses to information conveying Leo's false and true beliefs. We articulated Leo's beliefs through a single scenario where the positions of the two containers — namely, the box and the bucket — remained constant, and the protagonist consistently remained present without any instances of absence or return. Consequently, we propose that Leo's belief task is conducive to identifying the neural correlates of false belief processing in toddlers.

A notable strength of our method is using independent tasks to assess false belief understanding on both neural and behavioral levels. The behavioral task, designed with low response-generation and inhibitory-control demands, enabled the categorization of toddlers into false belief passers and failers. Concurrently, the ERP belief task provided explicit information about reality and the protagonist's beliefs, eliciting toddlers' neural responses. The deliberate differentiation between the two tasks ensured that the identified brain-behavior connections reflected substantive cognitive processes in false belief understanding rather than superficial task characteristics.

In summary, the present study has two goals. The first goal is to explore late waveforms distinguishing false belief and true belief processing in 33-36-month-olds. Based on previous ERP studies on false belief understanding, we hypothesize that late waveforms over anterior or posterior electrode sites will differentiate between the false belief and true belief conditions[4]. The second goal is to determine if behavioral false belief competence is associated with a neural-level distinction. We measure behavioral false belief competence with the low-demands false belief task adapted from Setoh et al. (2016). If such competence aligns with a distinction on the neural level, we anticipate late waveforms to manifest exclusively in behavioral false belief passers, as

---

[4] Since it is unclear whether the term "slow waveform" can be applied to characterize late waveforms emerging as neural correlates of mentalizing in toddlers, we opt to speak more generally of a "late waveform" in the present ERP study of false belief reasoning in toddlers.

demonstrated by Liu et al. (2009b).

2.2. Study 1 methods

2.2.1. Participants

A total of 128 toddlers participated in the ERP belief task (55 boys, M_{age} = 34.82 months, *SD* = 1.71 months, age range: 32.50-37.37 months). Data obtained from 70 toddlers (30 boys, M_{age} = 34.84 months, *SD* = 1.71 months, age range: 32.50-37.23 months) fulfilled the artifact criterion, which will be explained in the subsequent section. These data constituted the final ERP dataset. Fifty-eight toddlers were tested but not included in the final sample because of the following reasons: 1) they did not achieve the requisite number of trials (a minimum of five usable trials, Elsner et al., 2013; Stets et al., 2012; Stets & Reid, 2011) due to challenges arising from inattention, excessive body movements, and tiredness (*n* = 49), 2) technical problems (e.g., no record stimuli markers in these toddlers) during the data recording (*n* = 6), or 3) the data recording process was not completed (*n* = 3). The exclusion rate was 41.18%, similar to other EEG studies with young children (Hoehl & Wahl, 2012; Stets et al., 2012).

Out of the 70 toddlers who met the ERP artifact criteria as mentioned earlier, 25 were excluded from the analysis of the behavioral false belief task for the following reasons: 1) Ambiguous or absent responses (*n* = 4); 2) Lack of engagement with the task (*n* = 15); 3) Errors in task instructions by experimenters (*n* = 6). Detailed coding schemes and reasons for exclusion were listed in the supplemental materials (S3). Forty-five (18 boys, M_{age} = 34.51 months, *SD* = 1.62 months, age range: 32.93-37.23 months) of the 70 toddlers who met the ERP artifact criteria and completed the low-demands behavioral false belief task concurrently were included in the final sample for data analysis.

According to parent reports, all toddlers had normal or corrected-to-normal visual acuity, normal hearing, no neurological disorders, or regular medication. German was the mother tongue or the primary language of all participants. Prior to the commencement of data collection, written informed consent was obtained from a parent or legal guardian for each participating child. The local ethics committee approved the study based on the ethical principles of the European Federation of Psychologists' Associations.

2.2.2. ERP paradigm and procedure

2.2.2.1. ERP belief task (Leo's belief task)

Leo's belief task is an ERP paradigm designed to assess false belief understanding in toddlers. Thirty-two test trials (16 false belief and true belief condition trials, respectively) and eight control conditions were presented to participants in a pseudo-randomized order. All auditory commentaries were recorded in high quality in WAV format (44.1KHz / 16bit sampling) by a trained speaker and played back at a volume of approximately 65 dB/A via JBL Control One monitor loudspeakers.

Five static scenes were presented in each trial as a film sequence with verbal commentary (**Figure 9**). The first scene depicted two containers, a box (German: Koffer) and a bucket (German: Eimer), which remained in the same positions throughout the trial. In the second scene, one of four possible objects (a ball, a fish, a cup, or a car) was shown in one container, and the following verbal comment was made: "Look! This is Leo's ball. The ball is in the box." (German: "Schau! Das ist Leo's Ball. Der Ball ist im Koffer."). Object positions were counterbalanced across all trials. The third scene showed the containers closing, accompanied by a corresponding sound. In the fourth scene, the protagonist, Leo, appeared on the screen and expressed his belief about the object's location ("Leo thinks the ball is in the box/bucket", German: "Leo denkt, der Ball ist im Koffer / Eimer."). In the true belief condition, Leo's belief corresponded to the object's true location. In the false belief condition, Leo's belief contradicted the real location of the object. In the control condition, Leo's belief was replaced by a location preference statement (e.g., “Leo likes the box/bucket”). Leo’s preferences were counterbalanced across control conditions. The final container word (i.e., box or bucket), signifying Leo’s presumed location for the object, was used to segment ERP events. The two containers (i.e., box and bucket) were shown again in the fifth scene. Two filler sentences were used to prevent the behavioral response from being influenced by the aforementioned location. The empty container was indicated (i.e., “The box/bucket is empty.” German: “Der Koffer/Eimer ist leer.”), and the object’s true location was repeated (e.g., “The ball is in the box.”, German: “Der Ball ist im Koffer.”). The order of these two filler sentences was counterbalanced across trials. Each trial ended with a test question (Belief conditions: “Where will Leo look for the ball?”; Control conditions: “Where will Leo go now?”). The final scene remained on the screen until the child responded orally.

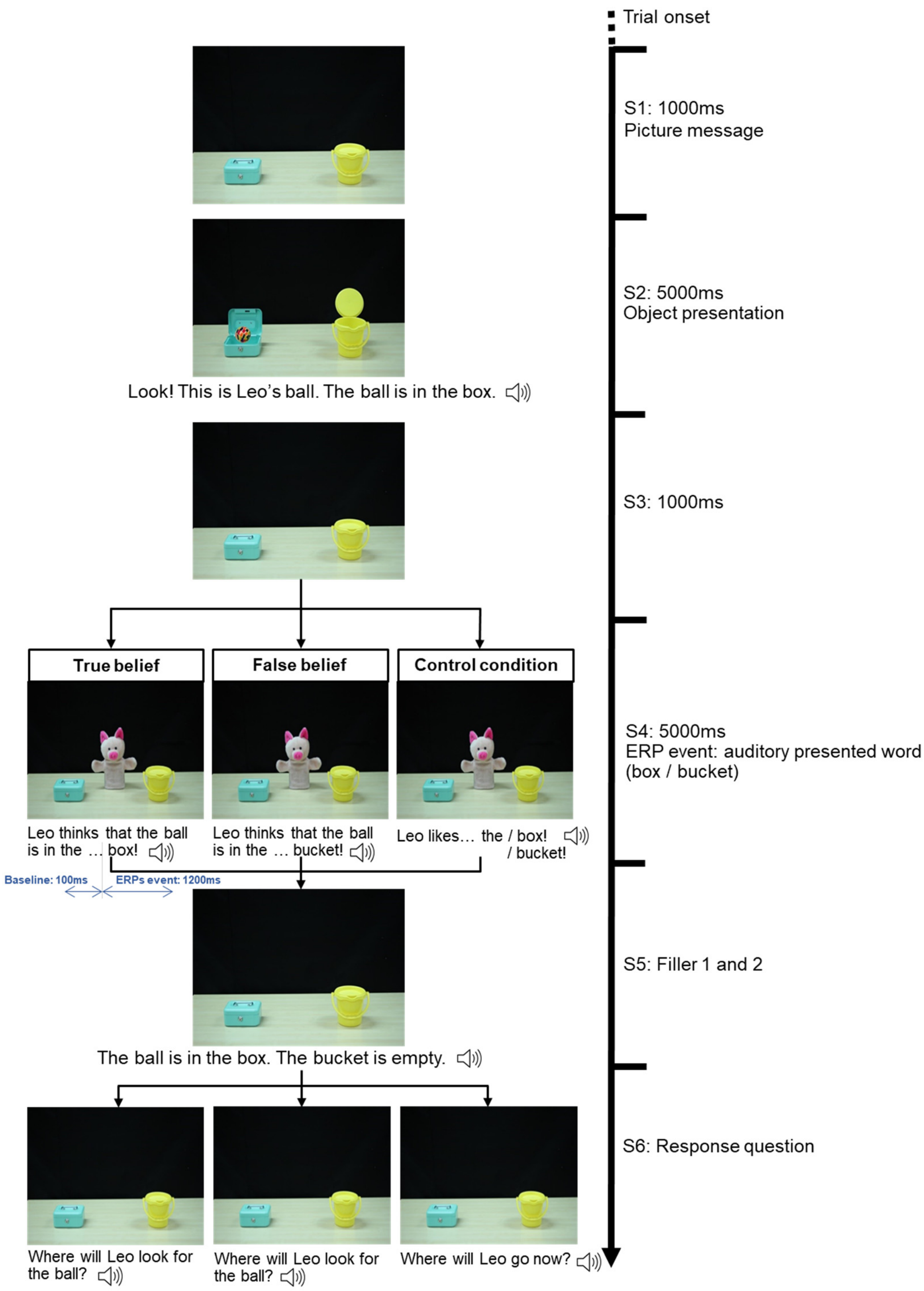


Figure 9. Illustration of ERP belief task. ERP events were elicited from the auditory presented word box/bucket in all experimental conditions. Control conditions were presented pseudo-randomly between the experimental conditions to prevent anticipation. The loudspeaker symbol indicates the auditory presented information.

2.2.2.2. Procedure

ERP belief task was conducted in a dimly lit, soundproof, and electromagnetically shielded room by IAC (Industrial Acoustics Company). Toddlers sat on a highchair in front of the experimenter or on their parent's lap if preferred. Stimuli were displayed on a 19-inch screen (Dell Inc.) with a resolution of 1280 × 1024. The image on screen measured 25.5 cm in height and 38 cm in width, with the protagonist occupying roughly two-thirds of the image on a dark grey frame background [RGB (96, 96, 96)], which was completely visible during inter-trial breaks.

At the start of the experiment, toddlers were asked to watch the following videos carefully and to answer the question at the end of each trial. Prior to the start of the trial, short "eye-catching" videos were interspersed to draw the toddlers' attention to the monitor and counteract disinterest. The entire task took about 15-20 minutes (including short breaks), plus another 25-30 minutes for preparation (application of the cap, further technical preparations, and instruction). Presentation software was used to deliver the stimulus (Neurobehavioral Systems).

2.2.2.3. Electrophysiological recordings and preprocessing

The electroencephalogram (EEG) was recorded with Cz reference from 33 active electrodes (ActiCap System, Brain Products, Gilching, Germany) placed on standard positions of the extended international 10–20 System. The bandpass of the recording system (BrainAmp DC amplifiers, Brain Products, Gilching, Germany) was set to 0.016 Hz-100 Hz, and data were sampled at 500 Hz. All impedances were kept below 10 kΩ. Fp2 was used to monitor vertical eye movements and blinks. Horizontal eye movements were monitored via positions F9 and F10. Channel AFz served as grounding.

The EEG data were analyzed offline using Brain Vision Analyzer software (Version 2.2.2, Brain Products GmbH, Gilching, Germany). Channels (i.e., F7, F8, T7, T8, P7, P8, TP9, TP10) near the EEG cap's edge (e.g., neck, cheeks, and forehead) were omitted since they tended to produce most of the movement-related noise. Data from EEG channels over the scalp surface (F3, F4, Fz, Fpz, FC1, FC2, FC5, FC6, C3, C4, Cz, CP1, CP2, CP5, CP6, P3, P4, Pz, O1, O2, Oz, n = 21) and EOG related electrodes (F9, F10, Fp1, Fp2) were retained for further analysis (see **Figure 10**). Offline data were re-referenced to the common average reference[5] and were digitally bandpass filtered from 0.30 Hz (-24 dB) to 30 Hz (-12 dB). An automatic inspection procedure was applied to discard data containing excessive eye movement or muscular artifacts before the ocular correction (Gratton et al., 1983). In this procedure, sections were automatically marked as artifacts and excluded if the amplitude of the EOG channel exceeded ± 125 μV and the EEG channel exceeded ± 100 μV. ERP segments time-locked to the ERP trigger (the onset of the auditory presented word, i.e., box, bucket) were extracted from -100 ms (pre-stimulus baseline) to 1200 ms after stimulus onset. Following the automatic artifact reduction process, a meticulous visual inspection was conducted. Video sequences (Analyzer Video Plug-In, Brain Products) recorded simultaneously with EEG data were used to identify visible inattentiveness (such as toddlers looking away from the monitor) and inappropriate activities, further mitigating the potential contamination from eye movements and other artifacts. Segment rejection was done blind to conditions by the experimenter. Detailed coding schemes for segment exclusion were listed in the supplemental materials (S2).

[5] Common average reference has been critiqued for possibly inducing 'mirror potentials,' where specific ERP waveforms might appear as their inverse at different scalp locations (Dien, 1998; Picton et al., 2000). To assess such possible bias, we reprocessed our data using the linked-mastoid re-reference, a technique less prone to mirror potentials (see supplementary materials S6). The late waveform patterns and their topographical characteristics were consistent irrespective of the re-referencing strategy, thereby ensuring the robustness of the observed polarity reversal in ERP patterns.

Subsequently, data were baseline corrected (−100 ms to 0 ms), and artifact-free segments were averaged separately for the false belief (*M* = 7.11, *SD* = 2.01, value range: 5 – 12) and true belief (*M* = 7.61, *SD* = 2.83, value range: 5 – 15) conditions. The number of usable trials did not differ significantly between groups ($F(1, 43) = 0.552$, $p = .461$), did not differ significantly between belief conditions ($F(1, 43) = 2.668$, $p = .110$), and did not interact between belief and group ($F(1, 43) = 0.830$, $p = .367$).

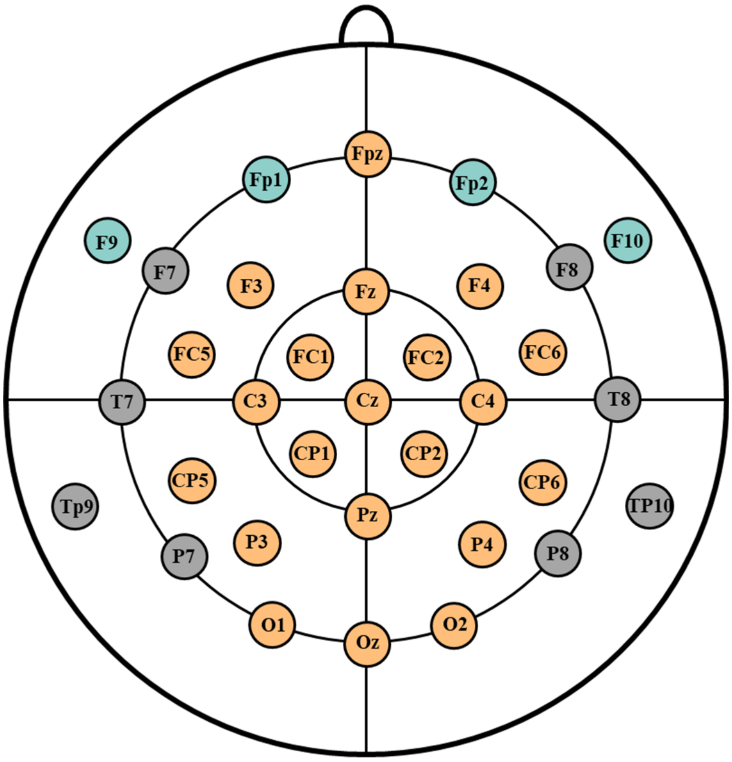


Figure 10. Electrode layout of the ERP measurement. All channels were included in the ERP measurement. Channels marked in teal and orange show the electrodes included in the ERP preprocessing.

2.2.3. Behavioral tasks and procedure

2.2.3.1. Low-demands behavioral false belief task (Lily's belief task)

Lily's belief task, which closely followed the procedure by Setoh et al. (2016), is a change-of-location false belief task with reduced processing demands. We created a picture book with nine pages to present the task, composed of clear plastic sheet protectors (32 cm × 56 cm) holding white paper backgrounds to which 11 pictures were attached (20 cm × 25 cm). A solid black paperboard (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-degree angle via a black ribbon connecting 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.

During the low-demands behavioral false belief task (**Figure 11**), the child and one experimenter sat beside each other with the easel in the middle of the table. The event was recorded with a hand-held camera focused on the book and the child, capturing the child's pointing movements. Another experimenter documented the child's behavior during the experiment. A verbal answer by the child and ambiguous response behavior (e.g., the pointing gesture and the verbal answer did not match) were recorded for all exercises and test runs. Six story events, two practice trials, and one test trial were presented to the child in total. After flipping each page to reveal the respective picture, the experimenter repeated the accompanying line of that event verbatim. In the first two pictures, 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 was shown, and the child was asked, “Where is Lily's apple?”. Then, in the third and fourth pictures, Lily moved the apple into the 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 picture, 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 the apple was located was coded as a 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.

All responses from the child were independently coded from video recordings by 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 vocal reply of the child and eventual response behavior (e.g., inconsistent verbal or pointing responses, such as pointing at one location but mentioning the other). The toddlers who completed the behavioral false belief task were categorized into two groups: those who passed were coded as 1 (passers), while the others were coded as 0 (failers).

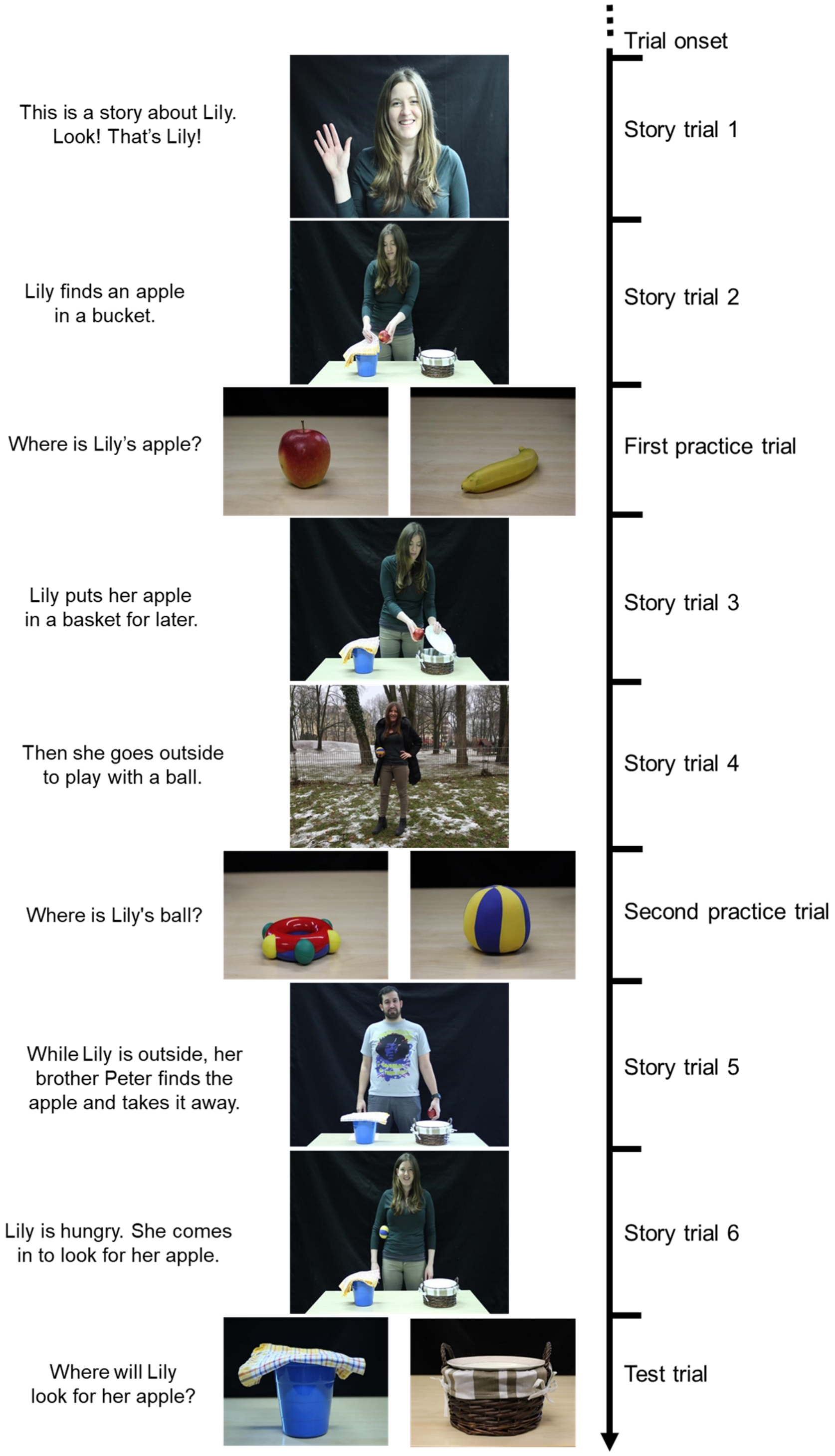


Figure 11. Illustration of the low-demands behavioral false belief task. The participant saw a six-part story with two practice questions and a test question. After each page turned, the experimenter read out the text on this page. A pointing gesture or verbal response to the container (basket) in which Lily left the apple counted as a correct answer.

2.2.3.2. Control measures

To control for the effects of general cognitive abilities and language skills, we conducted the Kaufman Assessment Battery for Children-Second Edition (KABC-II) (Melchers & Melchers, 2015) and the Language development test for three- to five-year-old children (SETK3) (Grimm et al., 2015).

In the KABC-II task, the toddlers underwent the five age-appropriate tasks, which collectively measure three-year-olds' intellectual processing index. These tasks assessed toddlers' memory, attention, and concentration, as well as their simultaneous and sequential information processing and coding.

We conducted four subtasks from the SETK3 test battery: comprehension of sentences, encoding of semantic relations, phonological working memory, and morphological rule formation. All tasks were conducted and scored as described in the manual, and the raw values were transformed into standardized T-values according to the age-specific norm table.

2.2.4. Statistical analysis

2.2.4.1. False belief competence grouped by false belief competence in the low-demands behavioral task

We used nonparametric cluster-based permutation analysis to determine the electrode clusters and time windows of interest for late waveforms. Specifically, we submitted EEG data in the time window from 600 to 1200 ms after the auditory presented word (i.e., box or bucket) and over all electrodes to a dependent-sample permutation t test to examine the main effect of the belief condition. The main effect of belief condition has been extensively demonstrated in both adult and child studies (e.g., Geangu et al., 2013; Meinhardt et al., 2011). This specific time window was selected based on prior studies in children, which consistently observed the emergence of late waveforms associated with belief reasoning approximately 600 milliseconds after stimulus onset (e.g., Bowman et al., 2012; Geangu et al., 2013; Meinhardt et al., 2011, 2012). Then, the mean amplitude of the late waveforms per belief condition was calculated over the electrode clusters, which revealed a significant difference between the false belief and true belief conditions. To confirm the significant effect of the belief condition and to explore the potential interaction between group, belief condition and electrode cluster, we performed a 3-way mixed-measures ANOVA on these data. In cases where a significant 3-way interaction was observed, we conducted 2 kinds of post-hoc analyses: (i) to further explore the effects of belief condition and electrode cluster on late waveforms' amplitude, we conducted post-hoc 2-way ANOVAs separately for the passers and failers, and post hoc pairwise comparisons were conducted when a significant 2-way interaction was observed. (ii) To further explore performance differences between the 2 groups in each sub-condition, we broke down the 3-way interaction by focusing on the effect of the group for each level of the other 2 factors (i.e., belief condition and electrode cluster). However, late waveforms' amplitude between the 2 groups was not further explored when neither the 3-way or 2-way interaction between the group and other factors nor the main effect of the group was significant.

2.2.4.2. False belief competence grouped by performance on the behavioral test questions in the ERP belief task

Due to generally low accuracy in the ERP belief task, classifying children into passers and failers directly based on performance in the behavioral test questions was not feasible. Instead, we used an independent low-demands behavioral task where children performed above chance levels, allowing for classification into false belief passers and failers. This design might raise concerns about whether the ERP task and the behavioral task reflect the same belief understanding process. To address this concern, we conducted an additional analysis focusing solely on the ERP task (Refer to section 3.2.2 for detailed results).

We used nonparametric cluster-based permutation analysis in all toddlers who fulfilled the ERP artifact criterion (N = 70) to determine the electrode clusters and time windows of interest for late waveforms. Specifically, we submitted EEG data in the time window from 600 to 1200 ms after the auditory presented word (i.e., box or bucket) and over all electrode clusters to a dependent-sample permutation t test. Then, the mean late waveform amplitude per belief condition and electrode clusters was calculated over the cluster, which revealed a significant difference between the false and true belief conditions. We then categorized children based on their performance in the ERP task into two groups: the top 25% for false belief accuracy (40% to 63.64%) and the bottom 25% for false belief accuracy (0% to 12.5%). This approach aimed to amplify differences despite overall low accuracy. To confirm the significant effect of belief condition and to explore the potential interaction between group, belief condition and electrode cluster, we performed a 3-way mixed-measures ANOVA on these data and conducted planned pairwise comparisons between FB and TB in toddlers with top 25% FB accuracy, as well as between FB and TB in toddlers with bottom 25% FB accuracy.

All statistical analyses except permutation analyses were performed using the SPSS software (version 29; IBM Corp, 2023) and Rstudio (Version: 2024.09.0+375; RStudio Team, 2024). Cohen’s d was calculated as the effect size for t-tests. The Greenhouse-Geisser correction was applied in case of violations of the sphericity assumption. For the sake of brevity, the uncorrected degrees of freedom were reported.

Permutation analyses were performed using the FieldTrip toolbox (Oostenveld et al., 2010). This type of analysis allows for statistical tests over entire data points while still controlling for multiple comparisons (Maris & Oostenveld, 2007). More specifically, for each permutation test used in the present study, adjacent spatiotemporal points for which t-values exceed a threshold were clustered (dependent *t*-test; two-tailed; cluster-defining threshold $p = .05$; iterations = 5000). The absolute sum of the *t*-values within each cluster was defined as the cluster's weight. This weight served as the sole criterion for determining the cluster's significance. Cluster-based permutation estimates the likelihood of each cluster's weight in the actual data compared to random permutations of the dataset. The *p*-value for each cluster is defined as the proportion of random iterations that resulted in a higher cluster weight. Clusters with $p < .05$ were considered significant. For each significant cluster, we report the cluster weight, *p*-value, and the corresponding electrode clusters and/or time window.

2.3. Study 1 results

2.3.1. Behavioral results

Based on the low-demands behavioral false belief task performance, 27 toddlers were assigned to the passers group, while 18 toddlers were in the failers group. There was no significant difference between passers and failers in gender, age, general cognitive abilities, and language skills (see **Table 6**).

In the ERP belief task, mean response accuracy in the false belief condition (behavioral test question) for the passers and failers was 31.37% (*SD* = 18.30) and 28.61% (*SD* = 21.57), respectively. There was no significant difference in false belief accuracy between the two groups ($t(43) = 1.169$, $p = .249$). Mean response accuracy in the true belief condition for the passers and failers was 43.24% (*SD* = 23.27) and 51.50% (*SD* = 23.08), respectively, and there was no significant difference in true belief accuracy between the two groups ($t(43) = -0.460$, $p = .648$). In both groups, the mean response accuracy in the false belief condition was below 50% (passers: $t(26) = –5.291$, $p < .001$; failers: $t(17) = –4.207$, $p = .001$) and was close to 50% (passers: $t(26) = –1.508$, $p = .144$; failers: $t(17) = 0.275$, $p = .786$) in the true belief condition.

Additionally, we examined the correlation between averaged late waveforms in the ERP belief task and false belief competence in the low-demands behavioral false belief task (see **Table 7**). Despite controlling for co-developing abilities (i.e., IQ, language skills), the significant correlation between the late waveforms over occipital sites and false belief competence persisted.

Table 6. The participants' demographic information

	All participants	Passers	Failers	χ^2 | t	p
N	45	27	18		
Gender (m:f)	18:27	9:18	9:9	1.25	.264
Age (months)	34.51 (1.62)	34.49 (1.63)	34.54 (1.64)	0.087	.931
Age range (months)	32.93-37.23	32.93-37.23	33.00-37.03	-	-
KABC-II (N =36)	98.67 (8.41)	98.57 (8.91)	98.80 (7.95)	0.079	.937
SETK3 (N =35)	53.52 (6.24)	53.68 (5.66)	53.31 (7.13)	-0.171	.865

Note. Standard deviations are presented in parentheses. IQ is measured with the KABC-II task. Language skills are measured with the SETK3 tasks.

Table 7. Correlation between averaged late waveforms in the ERP belief task and false belief competence in the low-demands behavioral false belief task.

	TB_FC	FB_FC	Diff_FC	TB_Occ	FB_Occ	Diff_Occ
FB competence in low-demands behavioral FB task (*N* = 45)	0.14	0.12	0.00	-0.17	0.31* [$r_{partial}$ = 0.32*, *N* = 35]	0.39** [$r_{partial}$ = 0.40*, *N* = 35]

Note. * $p < .05$, ** $p < .01$ (one-tailed). The partial correlations controlling IQ and language are presented in brackets. Sample sizes are presented in parentheses. IQ is measured with the KABC-II task. Language skills are measured with the SETK3 tasks. FB = false belief. TB = true belief. TB_FC = average amplitude of late waveforms in the true belief condition over the frontocentral electrode sites. FB_FC = average amplitude of late waveforms in the false belief condition over the frontocentral electrode sites. Diff_FC = difference between average amplitudes in true and false belief conditions over frontocentral electrode sites. TB_Occ = average amplitude of late waveforms in the true belief condition over the occipital electrode sites. FB_Occ = average amplitude of late waveforms in the false belief condition over the occipital electrode sites. Diff_Occ = difference between average amplitudes in true and false belief conditions over occipital electrode sites. The time window for calculating the average late waveform amplitudes was based on the main effect of the belief condition identified in the cluster-based permutation test described in Section 2.3.2.

2.3.2. ERP results

2.3.2.1. False belief competence grouped by false belief competence in the low-demands behavioral task

Among 45 participants who met the ERP artifact criteria and completed the low-demands behavioral false belief task concurrently, we used nonparametric cluster-based permutation tests to identify electrode clusters and time windows showing potential differences in late waveforms' amplitude across the two belief conditions (false belief and true belief). The permutation analysis revealed significant clusters for the effect of belief condition on late waveforms' amplitude (Negative cluster: electrodes: FC2, FC6, C3, C4; time window: ~691-910 ms and ~936-1004 ms after the auditory presented word; t_{weight} = -1048.17 and -286.40; p = .002 and .023. Positive cluster: electrodes: O1, Oz, O2; time window: ~721-914 ms and ~924-1064 ms after the auditory presented word; t_{weight} = 932.76 and 633.84; p = .003 and .007). See **Figure 12** and the supplementary materials (S4) for the distribution of these significant clusters.

Based on the amplitude of the late waveforms averaged separately over the negative and positive electrode clusters, the 3-way mixed-measures ANOVA revealed a significant interaction between group, belief condition, and electrode cluster ($F_{1,43}$ = 4.21, p = .046, η_p^2 = .09) (see **Table 8**). To further explore the 3-way interaction effect, we conducted post-hoc pairwise comparisons between belief condition and electrode cluster on participants' performance on the low-demands false belief task; we conducted post hoc 2-way ANOVAs separately for the passers and failers. These analyses revealed significant interactions between belief condition and electrode cluster only in passers (passers: $F_{1,26}$ = 28.62, p < .001, η_p^2 = .52; failers: $F_{1,17}$ = 2.86, p = .109, η_p^2 = .14). Post-hoc pairwise comparisons (see **Figure 14A**) showed that, among the passers, the false belief condition triggered a significantly more negative amplitude than the true belief condition over the frontocentral electrode sites (t_{43} = 3.45, p = .001, Cohen's d = 0.38). The false belief condition triggered a significantly more positive amplitude than the true belief condition over the occipital electrode sites (t_{43} = -5.33, p < .001, Cohen's d = -1.05). Among the failers, the false belief condition triggered a more negative amplitude than the true belief condition over the frontocentral electrode sites (t_{43} = 3.45, p

= .001, Cohen's d = 0.38), while no significant amplitude difference between the false belief and true belief conditions was observed (t_{43} = -1.76, p = .467, Cohen's d = -0.18) over the occipital electrode sites. The waveforms and topographical maps elicited under different belief conditions and groups are shown in **Figure 13**.

Additionally, to explore amplitude differences between the 2 groups in each sub-condition, we also broke down the 3-way interaction by focusing on the effect of the group for each level of the other 2 factors (i.e., belief condition and electrode cluster). The results showed that amplitude was significantly more positive in passers than failers in the false belief condition over occipital electrode sites (t_{43} = -2.11, p = .041, Cohen's d = -0.61). However, there were no significant differences in amplitude between groups in the other 3 conditions (all $p's$ > 0.05; see **Figure 14B**). That is, within the context of participants who passed the low-demands false belief task compared to those who failed, the false belief condition elicited a higher amplitude over occipital electrode sites.

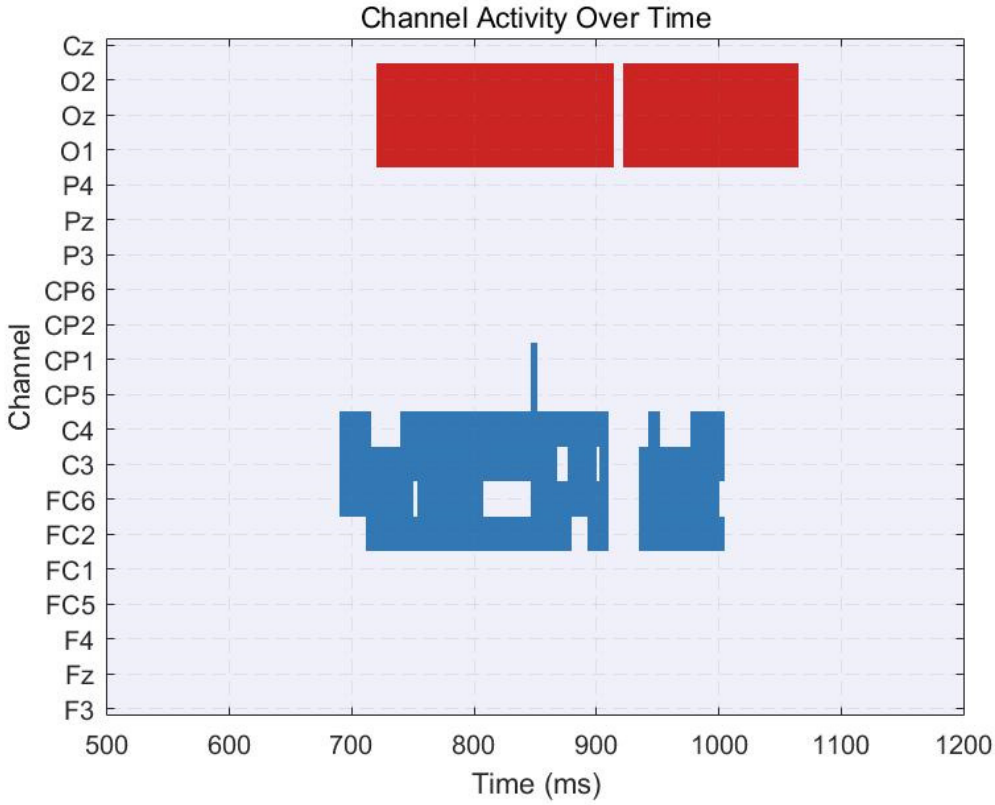


Figure 12. Raster diagrams depicting significant differences in amplitude between the false belief and true belief conditions from the ERP belief task (i.e., Leo's belief task), as determined by a cluster-based permutation test (Sample size *N* = 45). Red and blue areas represent electrodes/time points where late waveforms' amplitude in the false belief condition are more positive or more negative relative to the true belief condition, respectively. Black areas indicate electrodes/time points where no significant differences were observed. In the first cluster (red area), significant effects are localized to electrodes O1, Oz, and O2. In the second cluster (blue area), significant effects are primarily localized to electrodes FC2, FC6, C3, and C4.

Table 8. ANOVA results on late waveforms amplitude, false belief competence grouped by false belief competence in the low-demands behavioral task.

Factor	$F_{(df=1,43)}$	p	η_p^2
Group	2.24	.142	.05
Belief Condition	2.93	.094	.06
Electrode Cluster**	12.24	.001	.22
Group × Belief Condition **	10.27	.003	.19
Group × Electrode Cluster	0.03	.861	.00
Condition × Electrode Cluster ***	21.98	< .001	.34
Group × Belief Condition × Electrode Cluster*	4.21	.046	.09

Note: * p < .05, ** p < .01, *** p < .001.

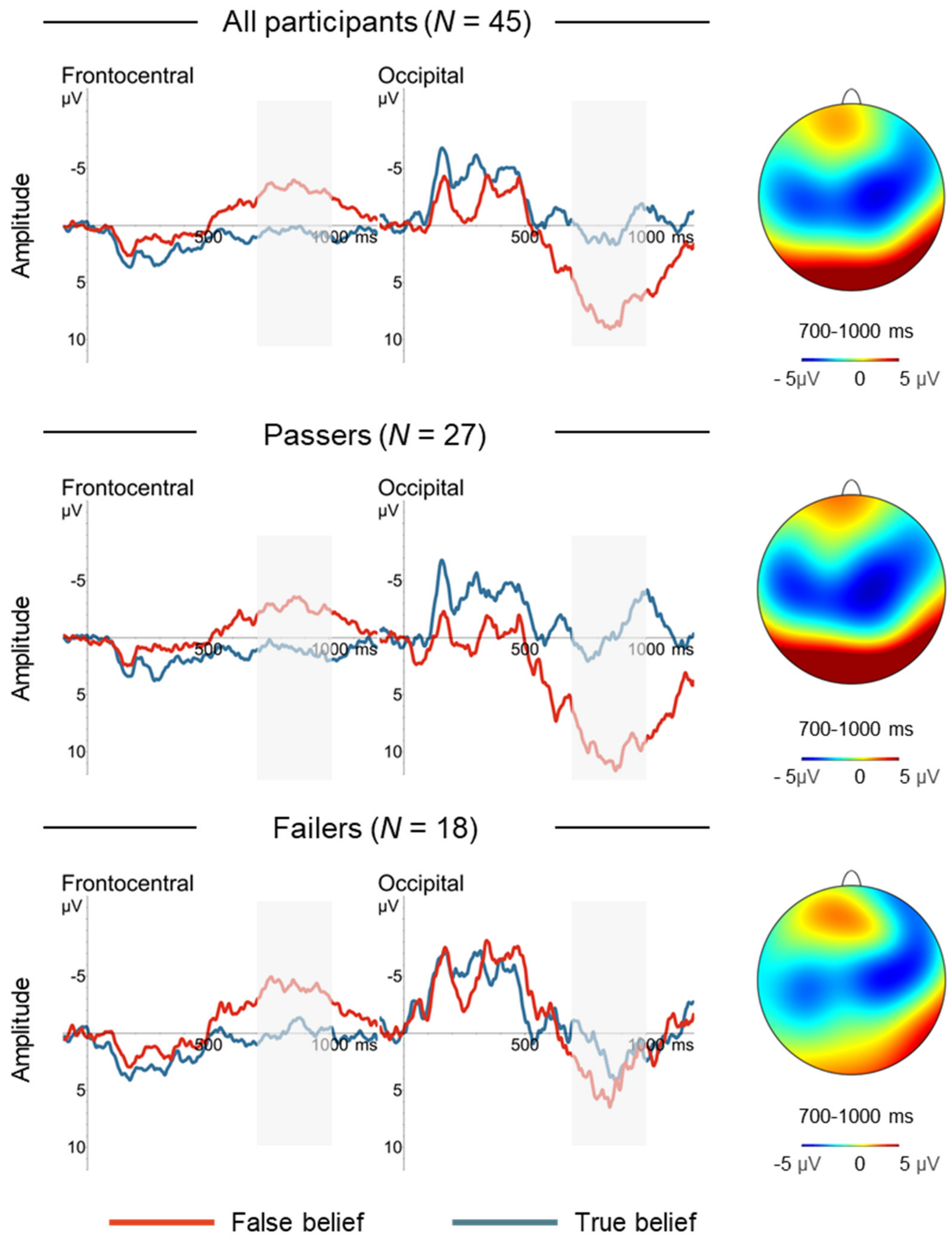


Figure 13. Late waveforms separated for "condition" (false belief, true belief) and "group" (passers, failers) plotted over the averaged frontocentral (across FC2, FC6, C3, C4) and occipital (across O1, Oz, O2) electrode clusters. Red line: late waveforms in the false belief condition; blue line: late waveforms in the true belief condition. Topographic maps showed the mean amplitude difference between late waveforms of the true belief condition subtracted from the false belief condition.

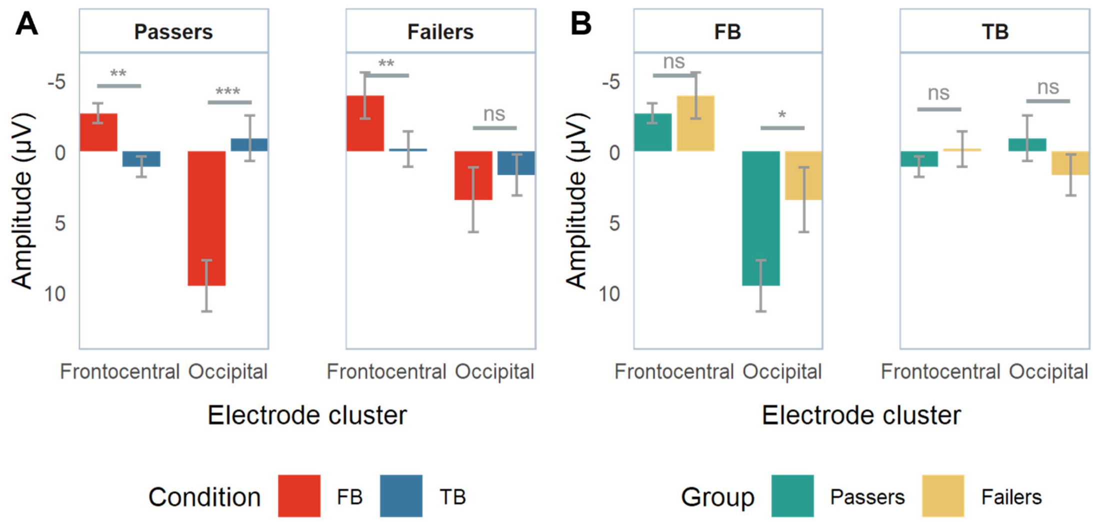


Figure 14. (A) A late positive waveform over the occipital electrode sites distinguished between the false belief and true belief conditions only in passers, while a late negative waveform over the frontocentral electrode sites consistently distinguished between the false belief and true belief conditions regardless of the low-demands behavioral false belief task performance. (B) Compared with failers, passers responded with a higher amplitude late waveform in the false belief condition over occipital electrode sites. Data are expressed as mean ± SE. Note that, in (A), the SE refers to the standard error of the pairwise difference between the 2 compared sub-conditions; in (B), the SE refers to the standard error of the difference between the 2 independent means. ns: not significant; * $p < .05$; ** $p < .01$; *** $p < .001$.

Furthermore, additional paired-sample *t* tests compared the false belief and control conditions over the occipital electrode sites (i.e., O1, Oz, O2) in separate groups to ensure that the late waveforms observed in the false belief condition in passers reflected false belief understanding rather than detecting a match or mismatch between Leo's beliefs and reality (see **Figure 15**). Among the passers, the false belief condition triggered a significantly more positive amplitude than the control condition (t_{23} = 3.09, p = .005, Cohen's d = 0.63) over the occipital electrode sites, while no significant difference between conditions was found among the failers (t_{14} = 0.86, p = .405, Cohen's d = 0.22).

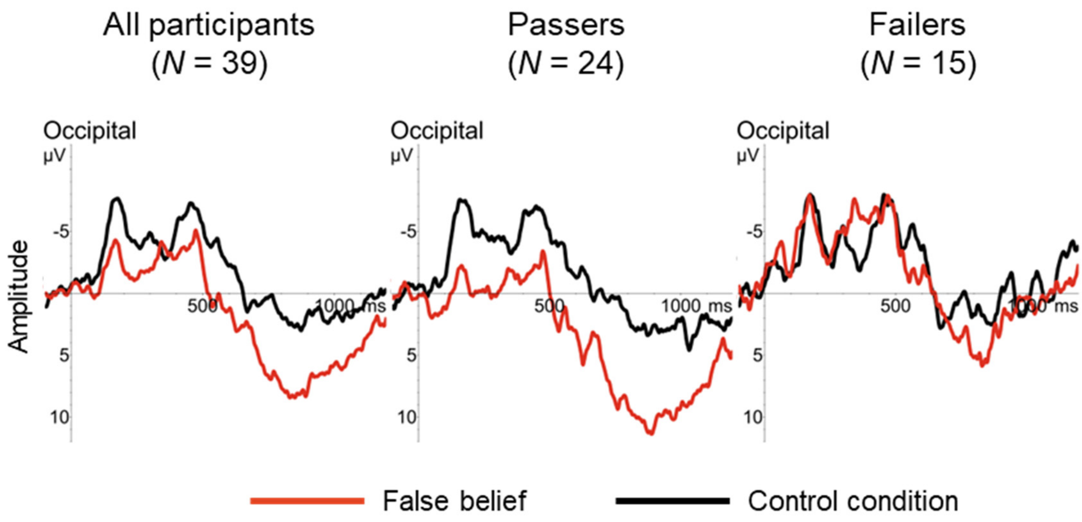


Figure 15. Late waveforms separated for "condition" (false belief, control condition) and "group" (all participants, passers, failers) plotted over the averaged occipital electrode sites (i.e., O1, Oz, O2). The additional analysis was narrowed down to 39 participants (Control condition: $M_{\text{-trials}}$ = 5.56, SD = 0.79, range: 5-8), as six others had fewer than five control conditions recorded, which made their data incomplete.

2.3.2.2. False belief competence grouped by performance on the behavioral test questions in the ERP belief task

Among 70 participants, the permutation analysis revealed significant clusters for the effect of condition on late waveforms amplitude (Negative cluster: electrodes: FC2, FC6, C3, C4; time window: ~625-807 ms and ~813-916 ms after the auditory presented word; t_{weight} = −884.37 and -477.75; p = .003 and .009. Positive cluster: electrodes: O1, Oz, O2; time window: ~607-1038 ms after the auditory presented word; t_{weight} = 1937.29; p < .001). See

Figure 16 and the supplementary materials (S5) for the distribution of these significant clusters.

Behavioral competence within the ERP belief paradigm, on average, was observed to be below chance levels, as anticipated. Consequently, to further investigate the brain-behavior connection, comparisons were made between extreme groups, specifically the top 25% (N = 18) and bottom 25% (N = 18) of performers based on false belief accuracy of behavioral test questions in the ERP belief paradigm. Among 36 participants, the permutation analysis revealed significant clusters for the effect of condition on the late waveform's amplitude (Negative cluster: electrodes: FC6, C3, C4; time window: ~653-747 ms and ~761-803 ms after the auditory presented word; t_{weight} = −458.74 and -204.24; p = .009 and .045. Positive cluster: electrodes: O1, Oz, O2; time window: 717-867 ms after the auditory presented word; t_{weight} = 572.81; p = .007).

Based on the amplitude of the late waveforms averaged separately over the negative and positive clusters, the 3-way mixed-measures ANOVA revealed a marginally significant interaction between group, condition, and electrode ($F_{1,43}$ = 3.19, p = .083, η_p^2 = .09) (see **Table 9**). To further explore the 3-way interaction effect, we conducted post-hoc pairwise comparisons between condition and electrode on participants' performance on the behavioral test questions in the ERP belief task; we conducted post-hoc 2-way ANOVAs separately for the top 25% group and bottom 25% group. These analyses revealed significant interactions between belief condition and electrode only in the top 25% group (top 25% group: $F_{1,17}$ = 21.70, p < .001, η_p^2 = .56; bottom 25% group: $F_{1,17}$ = 3.62, p = .074, η_p^2 = .14). Post hoc pairwise comparisons (see **Figure 18A**) showed that, among the passers, the false belief condition triggered a significantly more negative amplitude than the true belief condition over the frontocentral electrode sites (t_{34} = 4.56, p < .001, Cohen's d = 0.43).

The false belief condition triggered a significantly more positive amplitude than the true belief condition over the occipital electrode sites (t_{43} = -3.56, p = .001, Cohen's d = -0.89). Among the failers, the false belief condition triggered a more negative amplitude than the true belief condition over the frontocentral electrode sites (t_{34} = 3.86, p < .001, Cohen's d = 0.37), while no significant amplitude difference between the false belief and true belief conditions was observed (t_{34} = -0.85, p = .401, Cohen's d = -0.21) over the occipital electrode sites. The waveforms and topographical maps elicited under different belief conditions and groups are shown in **Figure 17**.

Additionally, to explore amplitude differences between the 2 groups in each of the sub-conditions, we also broke down the 3-way interaction by focusing on the effect of the group for each level of the other 2 factors (i.e., belief condition and electrode). However, there were no significant differences in amplitude between groups in all four conditions (all *p's* > 0.05; see **Figure 18B**).

The above findings are similar to those from the independent behavioral false belief task, enhancing the evidence on neural correlates of false belief understanding in toddlers. However, it is important to note that these findings should be viewed as supplementary due to the overall low accuracy in the behavioral test questions of the ERP task and the smaller sample sizes from the top and bottom 25%. The low-demands behavioral false belief task remains essential, as it offers a clearer performance distinction for grouping toddlers younger than 3 years into false belief passers and failers.

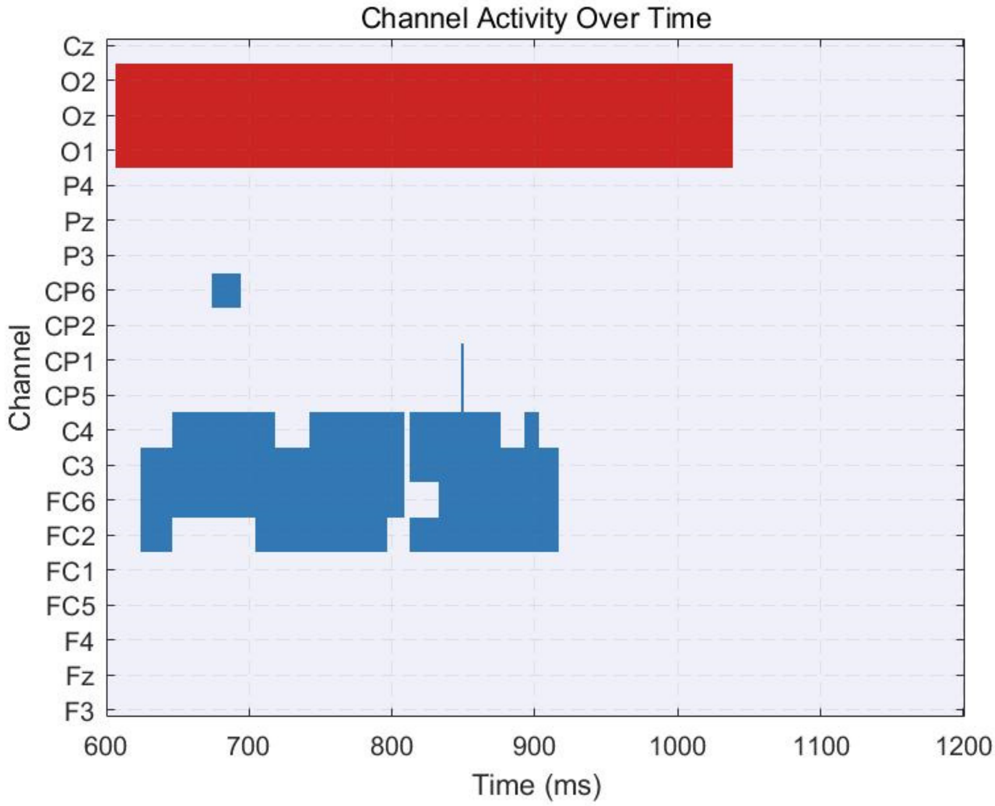


Figure 16. Raster diagrams depicting significant differences in ERPs between the false belief and true belief conditions from Leo's belief task, as determined by a cluster-based permutation test (Sample size *N* = 70). Red and blue areas represent electrodes/time points where ERPs in the false belief condition are more positive or more negative relative to the true belief condition, respectively. Black areas indicate electrodes/time points where no significant differences were observed. In the first cluster (red area), significant effects are localized to electrodes O1, Oz, and O2. In the second cluster (blue area), significant effects are primarily localized to electrodes FC2, FC6, C3, and C4.

Table 9. ANOVA results on late waveforms amplitude, and false belief competence grouped by performance on the behavioral test questions in the ERP belief task.

Factor	$F_{(df=1,34)}$	p	η_p^2
Group	0.30	.587	.01
Belief Condition	0.81	.375	.02
Electrode Cluster **	10.44	.003	.24
Group × Belief Condition †	3.31	.078	.09
Group × Electrode Cluster	0.64	.431	.02
Condition × Electrode Cluster ***	20.88	< .001	.38
Group × Belief Condition × Electrode Cluster †	3.19	.083	.09

Note: † $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$.

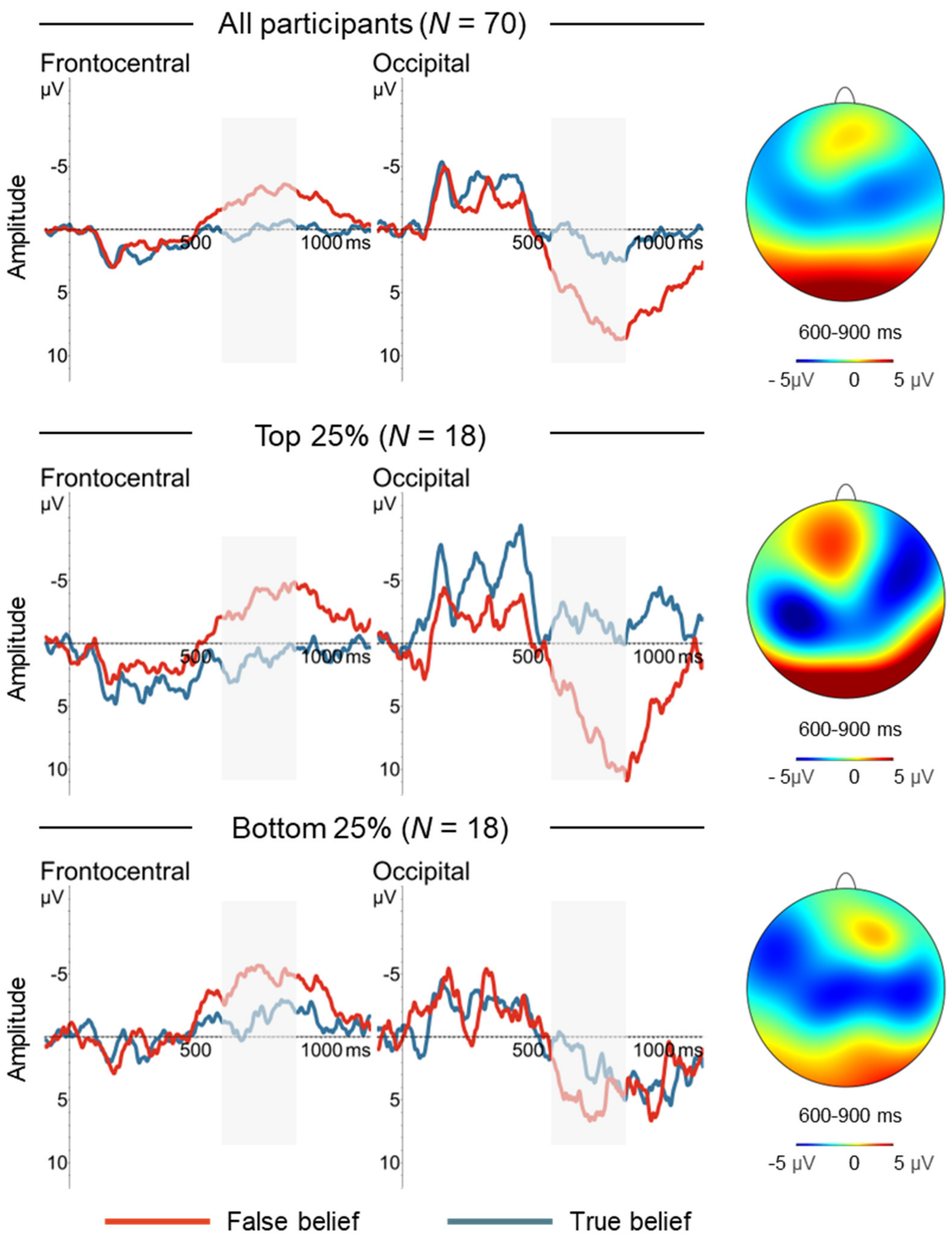


Figure 17. ERP waveforms separated for "condition" and "group" (children with top 25% and bottom 25% false belief accuracy in the ERP belief task) plotted over the averaged frontocentral and occipital electrode sites. Red line: ERP waveforms in the false belief condition; blue line: ERP waveforms in the true belief condition. Topographic maps showed mean amplitude differences between ERPs of the true belief condition subtracted from the false belief condition.

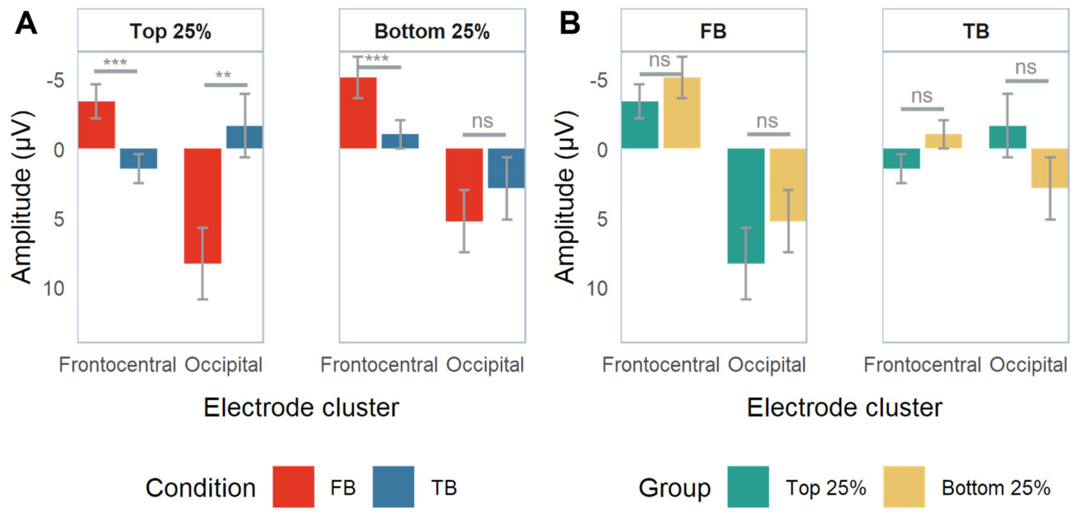


Figure 18. (A) A late positive waveform over the occipital electrode sites distinguished between the false belief and true belief conditions only in children with top 25% false belief accuracy in the ERP belief task, while a late negative waveform over the frontocentral electrode sites consistently distinguished between the false belief and true belief conditions regardless of false belief accuracy in the ERP belief task. (B) Compared with children with the bottom 25% false belief accuracy in the ERP belief task, children with the top 25% false belief accuracy in the ERP belief task responded with a higher amplitude late waveform in the false belief condition over occipital electrode sites. Data are expressed as mean ± SE. Note that, in (A), the SE refers to the standard error of the pairwise difference between the 2 compared sub-conditions; in (B), the SE refers to the standard error of the difference between the 2 independent means. ns: not significant; * $p < .05$; ** $p < .01$; *** $p < .001$.

2.4. Discussion on Study 1

The present study examined the neural correlates of explicit false belief understanding in toddlers aged 33 to 36 months. Recent behavioral research has shown that toddlers younger than 3 years can succeed in low-demands explicit false belief tasks, suggesting conceptual continuity in false belief understanding (Grosso et al., 2019; Scott et al., 2020; Setoh et al., 2016). However, to date, no research has investigated the neural correlates of explicit verbal false belief understanding in this age group. The present study found that early competence in a low-demands false belief task is paralleled by distinctive neural response patterns associated with false belief processing.

Our study revealed that toddlers aged 33-36 months exhibit distinct late waveforms in the false belief and true belief conditions across frontocentral and occipital electrode sites. Despite the absence of an inclusive description for the waveform that directly connects our findings with other studies, we compared the waveforms in our study and those that emerged in related studies. We found similarities with the late slow waveform observed in mentalizing in older children and adults in terms of late-onset and less distinct peaks, with distributions across frontal and posterior sites. Thus, the late waveforms elicited by false belief stimuli in toddlers appear to be akin to previous late waveforms, which studies linked to belief reasoning in adults and older children (e.g., Kühn-Popp et al., 2013; Liu et al., 2004, 2009b; Meinhardt et al., 2011, 2012; Sabbagh & Taylor, 2000). Nonetheless, the finding of a more diffused and less lateralized distribution of late waveforms in children who comprehend false belief is consistent with previous developmental cognitive neuroscience findings of a more diffused and bilateral brain activation in children, compared with adults performing the same cognitive tasks (Bowman et al., 2012; Kühn-Popp et al., 2013; Liu et al., 2004, 2009b, 2009a; Meinhardt et al., 2011, 2012). Notably, the brain-behavior correlations of false belief understanding were observed exclusively over the occipital electrode sites, consistent with prior research on older children's parieto-occipital ERPs involved in false belief reasoning (e.g., Guan et al., 2020; Kühn-Popp et al., 2013). The implications of these findings for understanding the neural underpinnings of belief processing in early childhood will be further discussed.

Following the approach by Liu et al. (2009b) in older children, this study grouped 33- to 36-month-old toddlers into passers and failers based on their performance in the low-demands behavioral false belief task by Setoh et al. (2016). A novel ERP paradigm was developed to examine the neural correlates of false belief and true belief reasoning in these groups. We identified neural responses over the frontocentral and occipital electrode sites that distinguished between the false belief and true belief conditions. For all participants, the amplitude differentiation of late waveforms was observed over the frontocentral electrode sites. However, the amplitude differentiation of late waveforms between the false belief and true belief conditions over the occipital electrode sites was observed only in passers, suggesting a connection to their behavioral false belief competence. This brain-behavior connection associated with false belief understanding expands on Liu et al.'s (2009b) findings, indicating a close relationship between behavioral false belief competence and a neural-level distinction between the false and true belief conditions. Our study directly contrasts the false belief condition with the true belief condition, underscoring the neural sensitivity of false belief reasoning over general mental state reasoning. Late waveform differences between false belief and control conditions further supported our hypothesis. This preliminary evidence suggests that late waveforms specifically reflect the neural processes underlying false belief understanding rather than broader cognitive mechanisms such as match or mismatch detection. It can be inferred that a neural system of belief-based processing is functional even before the third birthday.

While both Liu et al. (2009b) and the present study revealed neural sensitivity for false belief understanding in young children, they identified distinct brain regions engaged in belief reasoning. Specifically, we found engagement of occipital electrode sites in toddlers (i.e., 33-36-month-olds) with behavioral false belief competence, contrasting with Liu et al.'s (2009b) identification of frontal sites involvement in children aged 4 to 6 years with behavioral false belief understanding competence. These discrepancies may reflect developmental progressions in brain maturation but may also result from distinct effects of task configuration on neural responses.

Our findings, highlighting occipital late waveforms in toddlers with behavioral false belief understanding competence, differ from Liu et al.'s (2009b) identification of frontal late waveforms in older children. One possibility is that this difference may be due to the developmental maturation timeline. Specifically, the occipital regions mature earlier and undergo significant development in early childhood, while frontal regions show marked maturation after three years (Deoni et al., 2015). Such asynchronous maturation patterns hint at the posterior regions' earlier functional readiness for social cognition tasks, including processing belief-related information, compared to the anterior regions. However, given the cross-sectional nature of our comparison, which spans studies conducted in various countries with different tasks, coupled with the lack of precise localization in ERP studies, these interpretations should be considered preliminary. A nuanced understanding calls for further research to elucidate the developmental trajectory of neural mechanisms underlying false belief understanding and clarify how these mechanisms are refined with age.

Besides developmental brain maturation, task differences also impact the observed late waveform patterns between our study and Liu et al. (2009b). The two studies differed in ERP eliciting event configurations. In Liu et al. (2009b) study, the presentation of a story scenario was followed by a "think" and a "reality" question. The ERPs for the "think" and the "reality" conditions were time-locked to the "think" and "reality" events, which immediately followed the respective behavioral questions. Thus, Liu et al. (2009b) assessed the neural correlates of children's explicit verbal reasoning about true and false beliefs in comparison to reality. In contrast, our ERP eliciting events were activated by the auditory presentation of the last word of the belief information (e.g., "Leo thinks the ball is in the ***box***") without explicit belief reasoning instructions. Explicit belief reasoning instructions might engage neurocognitive processes differently than passive exposure to belief information, potentially tapping into a spontaneous representational understanding of beliefs (Aichhorn et al., 2009).

Our findings differed from Liu et al.'s (2009b) study in observing the false belief condition late negative waveforms over the frontocentral electrode sites, which appeared independent of behavioral false belief understanding competence. Compared to adults and older children (6-12 years) with behavioral false belief competence (e.g., Kühn-Popp et al., 2013; Liu et al., 2004, 2009b; Meinhardt et al., 2011, 2012; Sabbagh & Taylor, 2000), which also observed two distinct late waveforms, suggesting these ERPs as common neural responses in belief reasoning. Besides the late waveforms over the frontal sites, prior studies highlighted the presence of late waveforms over the parietal or parieto-occipital regions during belief reasoning tasks (Bowman et al., 2012; Guan et al., 2020; Kühn-Popp et al., 2013; Meinhardt et al., 2011). Our study corroborates this finding by detecting late positive waveforms in the posterior region, particularly towards the occipital electrode sites, reflecting broader and more delayed neural responses in toddlers compared to adults, supporting the notion of age-related changes in ERP latency (DeBoer et al., 2007; Taylor & Baldeweg, 2002) and the gradual enhancement of neural processing efficiency throughout childhood (Kail, 1991; Meinhardt et al., 2011).

Crucially, the association of only one of the two late waveforms with behavioral false belief competence in our study provides a nuanced perspective on their role in belief reasoning. Our results tentatively support a dual-process model for interpreting neural responses to false belief compared to true belief information. Specifically, the late negative waveforms independent of behavioral false belief understanding competence were observed over the frontocentral electrode sites. One interpretation of this result is that the frontal neural response primarily engages with the match or mismatch detection between Leo's belief and the state of reality rather than with the attribution of mental states. However, mismatch detection processes typically peak early, around 100 ms, reflecting a largely automatic process (Friederici et al., 2002; Nelson & McCleery, 2008). Since the present task required toddlers to process a complex sentence containing a sentential complement, a more delayed mismatch detection may be possible.

An alternative interpretation of the findings is that the late negative waveform over the frontocentral sites may be better compatible with the idea that two mentalistic response processes may be elicited almost simultaneously by the spontaneous response ERP task, a non-representational one over the frontocentral electrode sites and a representational one over the occipital electrode sites. False belief understanding involves a representational understanding of the mind, that is, an understanding that another person may misrepresent a state of the world, thereby seriously holding a claim to be true that the participant knows to be false. In early child development, non-representational but nonetheless mentalistic ways of handling falsity can also be observed. One prominent example is pretend play: The toddler who pretends to drink from an empty cup behaves as if there was tea in the cup but does not entertain this thought as a claim about the truth. A cognitive theory by Perner (1991, Chapter 8) distinguishes between "thinking of" (non-representational processing akin to desires or preferences) and "thinking that" (representational belief reasoning). The child who conceives of an agent as "thinking of" a state of the world (e.g., Leo thinks of the ball as being in the box) will recognize the discrepancy between Leo's thought and the real state of the world without representing Leo's thought as a misrepresentation of reality. Thus, the frontocentral neural responses to false belief versus true belief information observed in the present study may reflect the recognition of a discrepancy (as opposed to a correspondence) between a belief and a state of the world at a non-representational level. Support for this interpretation comes from an ERP study by Bowman et al. (2012), which found mid-frontal late waveforms related to both belief- and desire-reasoning in 7-8-year-old children, with distinct posterior late waveforms sensitive only to belief reasoning. This finding may indicate that false belief information was processed almost simultaneously at two levels, a non-representational and a representational level. The mid-frontal neural responses were common to belief- and desire-processing at a non-representational level.

In the present study, there is evidence for false belief reasoning over the occipital electrode sites. Late waveforms observed over the occipital electrode sites may appear to reflect participants' reasoning about beliefs as claims about the truth ("thinking that"). This kind of reasoning not only conceives of an agent as mentally connected to a state of the world but as correctly or falsely representing the state of the world. Thus, at this level, toddlers not only distinguish between mental states and states of the world but also understand the representational relation between them, which is equivalent to a representation of perspectives. This process of perspective representation is evidenced in the present study by late positive waveforms over the occipital electrode sites, which differentiated between the false belief and true belief conditions and were observed only in children who passed the behavioral false belief task. These results align with findings from Bowman et al. (2012), who identified selective posterior neural responses associated solely with belief reasoning in children who successfully answered the behavioral test questions. Similarly, contrasting false belief tasks with non-mental tasks involving misinformation (e.g., the false sign task) reveals sensitive belief representation in neural processing; Zhang et al. (2013) found that the false belief task elicited more pronounced positive waveforms in parietal sites compared to the false sign task. Moreover, neuroimaging studies, such as those by Aichhorn et al. (2009) and Perner et al. (2006), identified the right TPJ as a sensitive area for processing mental states and are selectively activated during false belief tasks compared with false sign tasks. Finally, recent research showed that the TPJ's involvement in false belief understanding becomes increasingly selective in younger age groups. A study employing the fNIRS technique (Wysocka et al., 2020) has shown that 3-5-year-olds exhibit sensitive activation patterns in the posterior regions during false belief reasoning tasks, reinforcing the notion of its critical role in the early development of Theory of Mind. Together, these findings pointed towards the sensitivity of posterior brain regions, including the TPJ, in false belief reasoning. However, the inherent limitations in the spatial resolution of ERP data preclude definitive identification of involved cortical structures. Further research is needed to dissect the complex neural mechanisms supporting false belief understanding.

Further neural-level comparisons between the passer and failer groups highlight distinct differences in the late waveform characteristics under both the false belief and true belief conditions over the occipital sites. These differences corroborate the observed brain-behavior correlations while suggesting potential patterns of cognitive resource allocation in processing belief information over the occipital electrode sites. Passers exhibited a more positive waveform over occipital sites in the false belief compared to the true belief condition, indicating greater cognitive resources allocation—such as attention and working memory—to the false belief condition. This likely reflects enhanced engagement in discerning between true and false scenarios. Conversely, in failers, there was no significant difference between the false belief and true belief conditions, as evidenced by the conflation of late waveforms across these scenarios. The amplitude of these subtle late waveforms, when compared to those observed in the false belief condition among the passers, suggests that the failers allocate a similar but constrained level of cognitive resources to both the false belief and true belief conditions. This pattern likely reflects failers' limited ability to effectively allocate cognitive resources and differentiate between false and true beliefs. These findings indicate that the development of false belief understanding abilities might be reflected in the allocation of cognitive resources during tasks involving belief processes.

Our study's limited number of trials may be seen as a limitation. In ERP research involving young children, there needs to be more certainty about the specific criteria governing the number or the proportion of trials necessary for stable ERP estimates. Previous work conducted a comprehensive review of ERP studies involving young children concerning infant ERP data for cognitive research, suggesting a minimum of five usable trials, with an average of eight trials for analysis (Brooker et al., 2020; Stets et al., 2012). Given the challenge of maintaining younger children's attention for extended ERP trials, our study emerges as a pioneering attempt to explore the neural correlates of false belief understanding in this population. Despite the relatively limited ERP trials (seven on average) in the false belief and true belief conditions in the ERP belief task, our study has yielded interpretable findings. Crucially, recent EEG research on infant social cognition (Filippi et al., 2020; Southgate & Vernetti, 2014) has demonstrated that meaningful conclusions about belief-based action prediction and neural correlates of infant action processing related to Theory of Mind can indeed be drawn from a similarly small number of artifact-free trials. This parallel in the literature emphasizes the potential utility of limited trial datasets in offering initial insights into early cognitive development. It cautiously suggests that our observations may contribute to a nascent understanding of the neural correlates of false belief understanding, highlighting the need for further research to solidify these early findings.

In summary, our study integrated a low-demands behavioral false belief task with a novel ERP paradigm to elucidate the brain-behavior connections of false belief understanding in children under three. Identifying one of two late waveforms associated with behavioral false belief competence supported a dual-process model of task performance, and an early sensitive neural system supporting false belief understanding. We observed late negative waveforms over the frontocentral electrode sites in both passers and failers, which may either reflect a match-mismatch detection between a proposition and a state of the world or the recognition of a discrepancy between a belief and a state of the world at a non-representational level. Crucially, late positive waveforms over the occipital electrode sites, observed exclusively in children who demonstrated behavioral false belief competence, appear to reflect a perspective representation process in toddlers. These findings suggest the possibility of early sensitivity in neural systems that support false belief understanding and thereby support the view that there is conceptual continuity in the development of false belief understanding. Further rigorous experimental and large sample-size studies are essential to validate and expand upon the current findings. Such studies will be crucial in refining our understanding of the cognitive and neural architecture of Theory of Mind from a developmental perspective.

Supplementary materials

S1. Seven ERP studies (Bowman et al., 2012; Kühn-Popp et al., 2013; Liu et al., 2004, 2009b, 2009a; Meinhardt et al., 2011, 2012) facilitated direct comparisons of neural responses associated with false belief understanding across adults and children by employing shared ERP paradigms. These studies are categorized into four groups, each using identical ERP paradigms in adults and children.

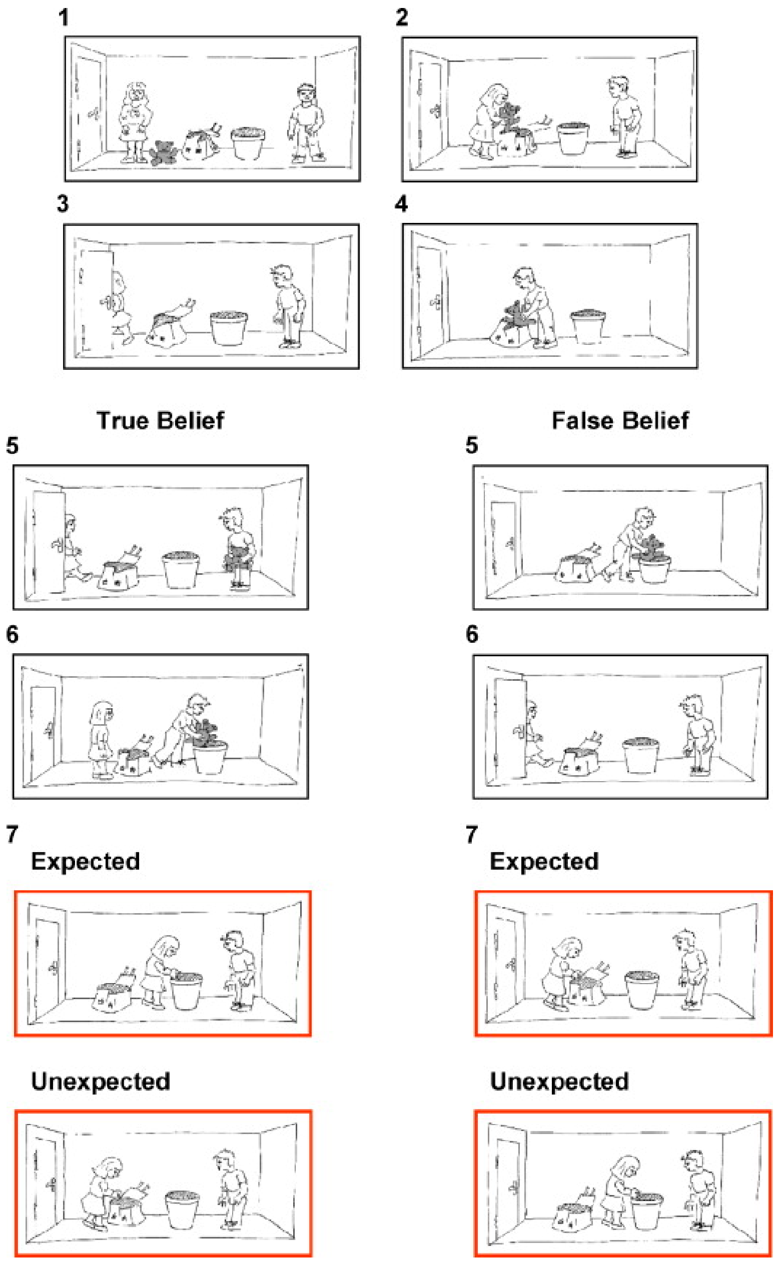


Figure S1. ERP paradigm from (Meinhardt et al., 2011).

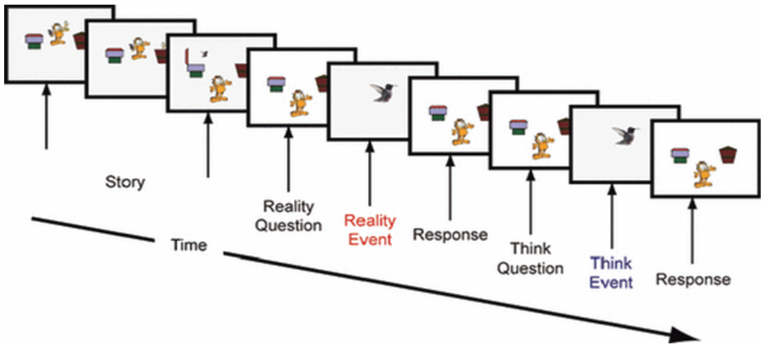


Figure S2. ERP paradigm from (Liu et al., 2004, 2009b)

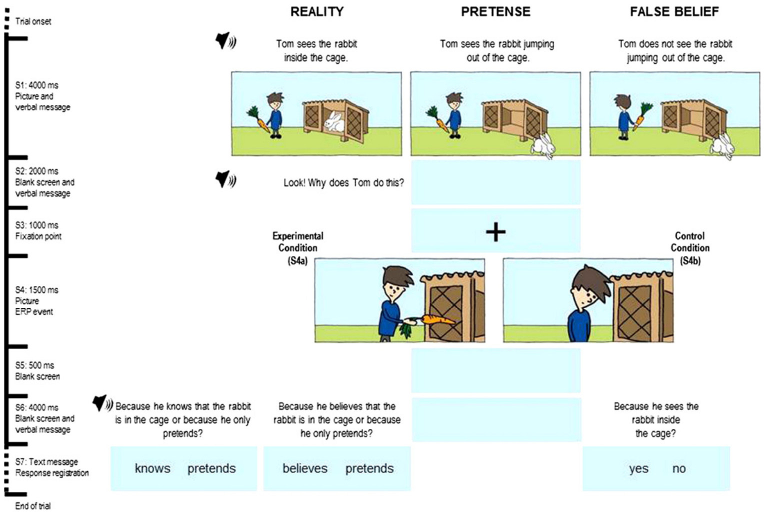


Figure S3. ERP paradigm from (Kühn-Popp et al., 2013; Meinhardt et al., 2012).

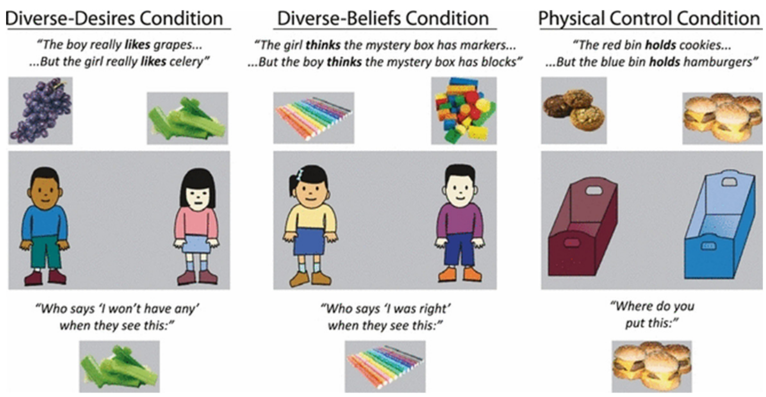


Figure S4. ERP paradigm from (Bowman et al., 2012; Liu et al., 2009a).

S2: The video sequences were evaluated by researcher experienced in EEG data preprocessing and familiar with EEG data collection protocols in children. Coders inspected each video for specific behaviors, notably instances where children looked away from the monitor or engaged in movements that could interfere with EEG data integrity. This visual inspection was conducted frame by frame to ensure no significant events were missed. To eliminate potential bias, coders were blind to the experimental conditions of the trials they reviewed.

Table S2. Video coding scheme for ERP belief task (i.e., Leo's belief task).

Participants
All children with ERP recording of Leo's belief task

Included segments	
Value	description
Included	Task engagement: Segments where the child maintains attention on the task presented on the monitor.
Excluded	1.Visual distraction: Children divert their gaze away from the monitor, breaking visual attention. 2.Physical movement: Segments featuring any physical movements, particularly of the head or upper body, which might disrupt EEG signal quality. 3.External interference: Parents or siblings are observed interfering with the child's engagement during the task. 4.Distraction from surroundings: Child was visibly distracted by other stimuli outside the task and does not focus on the task displayed on the monitor. 5.Lack of engagement: Child showed disinterest or an inability to continue with the task, such as persistent fidgeting or moving away from the setup.

S3. Detailed coding scheme of low-demands behavioral false belief task and the reasons and number of participants for their exclusion from further analysis in low-demands behavioral false belief task.

Table S3. Coding scheme of low-demands behavioral false belief task.

Inclusion or exclusion of Lily's belief task	
Value	Description
0	Not included in further analysis if 1) Child failed to answer practice questions. 2) Pointed incorrectly in one or both of the practice questions. 3) Experimenter asked "in the bucket or in the basket" or asked "here or here" or anything else pointing the two options. 4) Child replied with "at Peter" "it is gone" or similar. 5) Parents or siblings interfered. 6) Ambiguous reply (e.g., pointing at one thing but saying the other thing). 7) Experimenter made a different mistake. 8) Child did not reply or said I don't know. 9) Child did not conduct Lily's belief task.
1	Included in further analysis.
First practice question	
Value	Description
No value	Child fails to point at any of the pictures.
0	Child points at the picture of the banana.
1	Child points at the picture of the apple
Second practice question	
Value	Description
No value	Child fails to point at any of the pictures.
0	Child points at the picture of the rattle.
1	Child points at the picture of the ball.

Table S4. The reasons and number of participants for their exclusion from further analysis in low-demands behavioral false belief task.

Exclusion reasons	*N*
Experimenter asked “in the bucket or in the basket” or asked. “Here or here” or anything else pointing the two options.	6
Child replied with “at Peter” “it is gone” or similar.	1
Ambiguous reply (e.g., pointing at one thing but saying the other thing).	1
Child did not reply or said I don’t know.	2
Child did not conduct Lily’s belief task.	15

Note. N = number of participants.

S4. The distribution of significant clusters between the false belief and true belief condition in children who participated both ERP belief task (i.e., Leo's belief task) and low-demands behavioral false belief task (i.e., Lily's belief task) (N = 45).

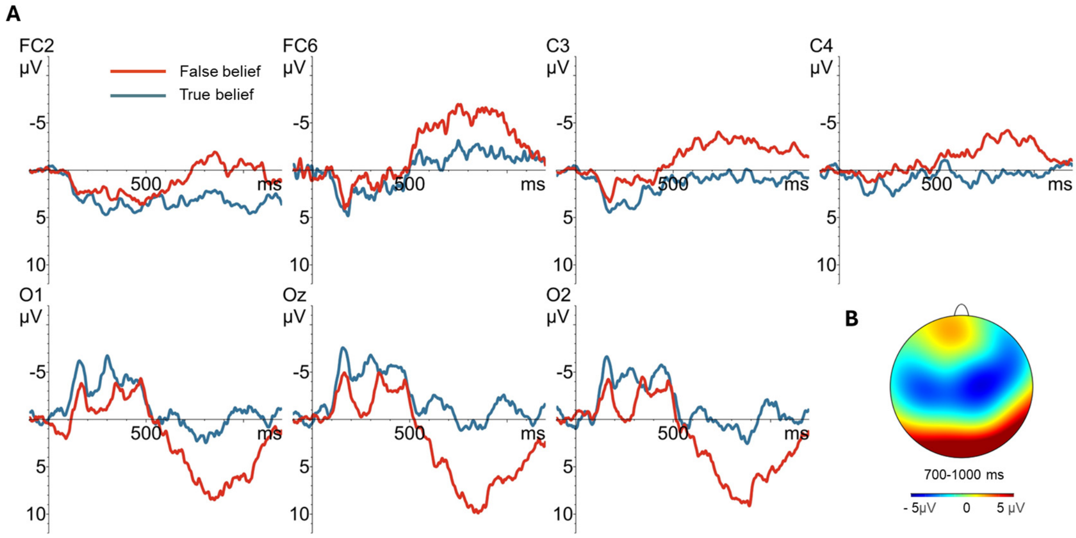


Figure S5. ERP data from all toddlers (*N* = 45) who completed ERP belief task and low-demands behavioral false belief task concurrently. (A) Grand average late waveforms in the true and false belief condition over the frontocentral and occipital electrode sites. (B) Topographic maps of the condition effect: mean amplitude difference between late waveforms of the true belief condition subtracted from the false belief condition.

S5. The distribution of significant clusters between the false belief and true belief condition in children with ERP belief task (i.e., Leo's belief task) (N = 70).

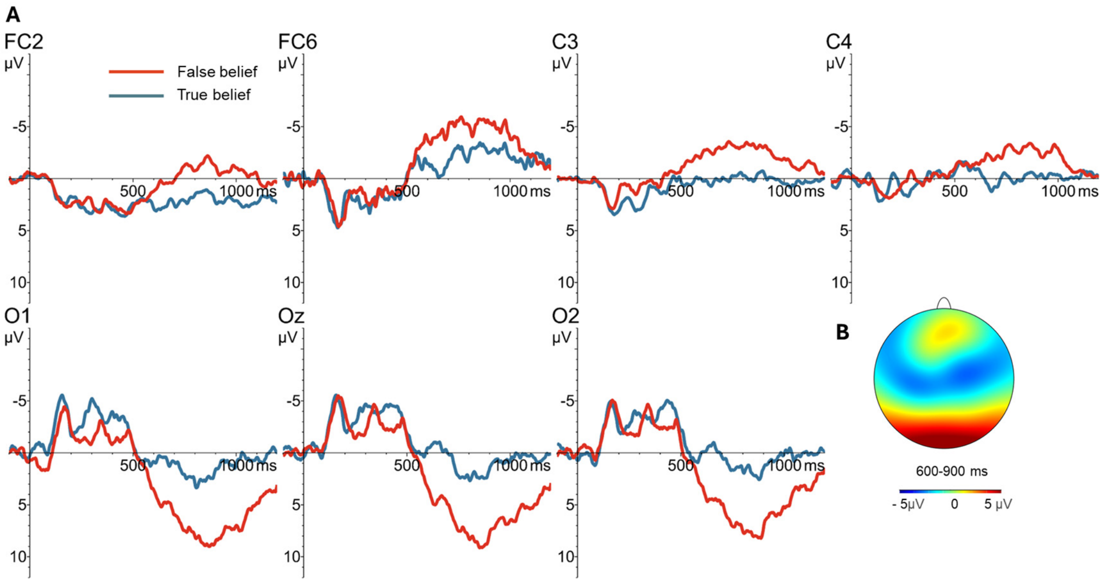


Figure S6. ERP data from all toddlers (*N* = 70) who fulfilled the ERP artifact criterion. (A) Grand average late waveforms in the true and false belief condition over the frontocentral and occipital electrode sites. (B) Topographic maps of the condition effect: mean amplitude difference between late waveforms of the true belief condition subtracted from the false belief condition.

S6. Common average reference has been critiqued for possibly inducing 'mirror potentials,' where specific ERPs might appear as their inverse at different scalp locations (Dien, 1998; Picton et al., 2000). To assess such possible bias, we reprocessed our data using the linked-mastoid re-reference, a technique less prone to mirror potentials. The late waveform patterns and their topographical characteristics were consistent irrespective of the re-referencing strategy, thereby ensuring the robustness of the observed polarity reversal in late waveform patterns. The late waveform patterns and their topographical characteristics are presented below. Except for the application of linked-mastoid re-referencing, all other preprocessing steps were identical to those described in section 2.2.2.3.

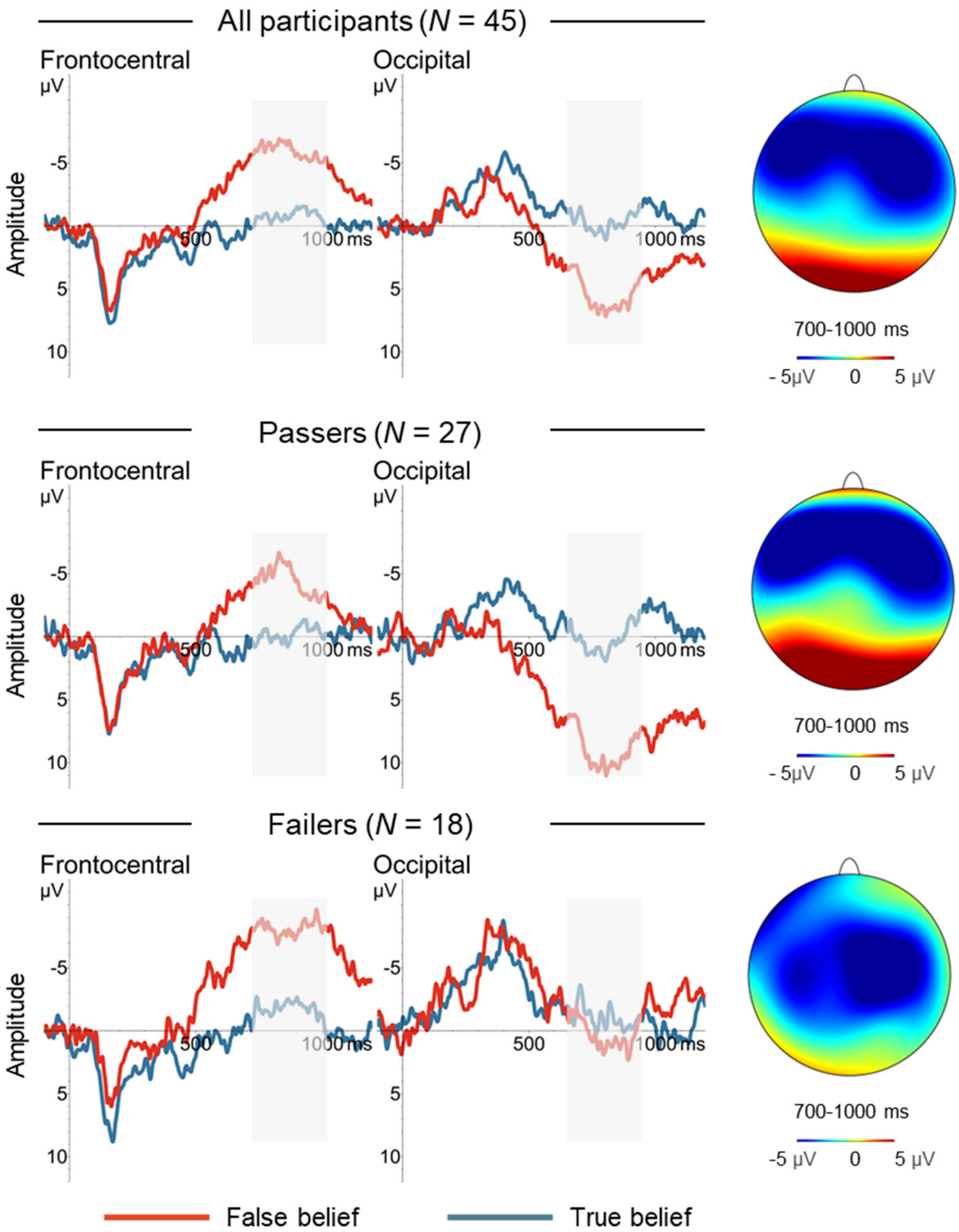


Figure S7. Late waveforms separated for "condition" (false belief, true belief) and "group" (passers, failers) plotted over the averaged frontocentral (across FC2, FC6, C3, C4) and occipital (across O1, Oz, O2) electrode sites (Sample size *N* = 45, re-reference in preprocessing: linked mastoid). Red line: late waveforms in the false belief condition; blue line: late waveforms in the true belief condition. Topographic maps showed mean amplitude difference between late waveforms of the true belief condition subtracted from the false belief condition.

Chapter 3: Resting-state EEG Alpha Asymmetry Predicts False Belief Understanding during Early Childhood: An Exploratory Longitudinal Study (Studies 2 & 3)

Highlights

1. This exploratory study demonstrated a correlation between resting-state EEG alpha asymmetry and false belief understanding during early childhood.
2. The present research utilized two independent longitudinal datasets that contained both EEG assessments (at 14 months and 34 months) and behavioral assessments of false belief understanding (at 34 and 52 months).
3. Superior explicit false belief understanding was associated with greater right relative than left frontal activity, while better implicit false belief understanding correlated with greater right than left parietal activity.
4. Resting-state EEG alpha asymmetry in frontal regions may serve as an early-appearing origin or precursor of children's later explicit false belief understanding.

Abstract

Theory of mind, the ability to attribute mental states to others, is fundamental to human cognition. In child development, a full or explicit understanding of false beliefs and their impact on action emerges around the age of 4 years. There is evidence of functional specialization of right hemispheric activity related to false belief processing in adults and children. However, it remains unclear whether this specialization is the cause or the consequence of Theory of Mind development. The present exploratory study investigates the longitudinal relation of resting-state electroencephalogram (EEG) alpha asymmetry measured in infancy/toddlerhood and behavioral false belief understanding at the age of 4 years. Employing a longitudinal design, Study 2 assessed resting-state EEG alpha asymmetry across frontal and parietal electrode sites at 34 months ($N = 43$), explicit false belief understanding at age 4 ($N = 22$), and implicit false belief understanding at both times (34 months: $N = 38$; 52 months: $N = 25$). Study 3 is another independent longitudinal dataset that included resting-state EEG alpha asymmetry at 14 months ($N = 37$) and explicit false belief understanding at age 4 ($N = 32$). We found that superior explicit false belief understanding at age 4 was associated with greater right versus left frontal activity at an earlier age, and better implicit false belief understanding was linked to greater relative right parietal activity. Given the limited sample size, these results should be viewed as preliminary and warrant replication in future studies. These findings, when interpreted cautiously, suggest that early asymmetric brain activity over frontal sites measured by resting-state EEG in children may predict their later explicit false belief understanding. Additionally, our results tentatively suggest that implicit and explicit false belief understanding may develop based on partly distinct neural mechanisms.

Keywords

False belief understanding, Theory of Mind, resting-state EEG alpha asymmetry, longitudinal study, toddlers.

Dataset source and Shuting Li's contributions to the study

Li, S., Müller, B. C. N., Meinhardt, J., & Sodian, B. (2025). Resting-state EEG alpha asymmetry predicts false belief understanding during early childhood: An exploratory longitudinal study. *Brain Research*, 149523. https://doi.org/10.1016/j.brainres.2025.149523

The data for Study 2 were collected as part of the project *"The Role of Language in Early Theory of Mind Development"* (DFG SO 213/33-1,2). The data for Study 3 were obtained from an earlier multi-measure longitudinal project, *"Theory of Mind in Infancy and Early Childhood"* (DFG SO 213/27,1-3). Shuting Li played a leading intellectual role in conceptualizing the paper, conducting data analysis, visualizing the results, and writing the manuscript.

3.1. Introduction

Theory of Mind is conceptualized as the ability to infer and understand the mental states of others, such as beliefs, desires, and intentions (Perner, 1991). A key entailment of a fully developed Theory of Mind is an understanding of the representational relation between mind and world, for instance, understanding that another person holds a false representation of a state of the world. Research indicates that a representational Theory of Mind, characterized by a full or explicit false belief understanding and related concepts, emerges in child development between the ages of 3 and 5 years (see Wellman et al., 2001, for a review). Recent evidence indicates an explicit false belief understanding in tasks with reduced processing demands in toddlerhood and an implicit false belief understanding in infancy (see Scott et al., 2022, for a review).

Existing studies on the neural correlates of false belief understanding in children predominantly focused on children aged 6 to 12 years with well-developed false belief understanding (e.g., Bowman et al., 2019; Gweon et al., 2012; Saxe et al., 2009). For example, a longitudinal study examined whether source-localized resting-state electroencephalogram (EEG) activity in the dorsal medial prefrontal cortex (MPFC) or the right temporoparietal junction (TPJ) at age 4 could predict ToM-specific functional magnetic resonance imaging (fMRI) responses 3.5 years later. The findings revealed that preschoolers' resting-state EEG activity in the dorsal MPFC predicted subsequent ToM-specific fMRI responses in the same region (Bowman et al., 2019). Additionally, in an event-related potential (ERP) study, Bowman et al. (2012) identified right mid-frontal activations associated with mental-state reasoning in children aged 7-8. Further fMRI studies with school-aged children (ages 5 to 11) suggest that the right TPJ increasingly distinguishes mental states from physical facts as they age (Gweon et al., 2012; Saxe et al., 2009). In contrast, there is a lack of research on the neural correlates of false belief understanding during the sensitive developmental period of 3 to 5 years. Grosse Wiesmann et al. (2017b) utilized structural MRI to show that false belief competence in 3- and 4-year-olds correlates with age-related increases in local white matter structure in regions such as the right ventral MPFC, right TPJ, and right PC. Moreover, a submitted ERP study in 33- to 36-month-old children indicated that late waveforms differentiated between the false and true beliefs, observed over slightly more right-lateralized frontocentral sites (see Study 1). Additional evidence supporting the right hemispheric lateralization of brain activity associated with false belief understanding comes from a resting-state EEG source-localized analysis by Sabbagh et al. (2009). They identified individual differences in alpha oscillations linked to Theory of Mind performance in 4-year-old children. Specifically, regions in the dorsal MPFC and several right hemisphere areas—including the TPJ, precentral gyrus, cuneus, and inferior temporal cortex—were associated with representational Theory of Mind, suggesting that the maturation of these regions may underpin the development of false belief understanding and related Theory of Mind abilities. Consistent findings were observed in adult populations. Sabbagh and Flynn (2006) found that greater activation (reflected by reduced alpha power) in the right mid-frontal region predicted better mental-state decoding performance, a component of Theory of Mind. This finding suggests a potential trait-like aspect of resting-state EEG alpha asymmetry in Theory of Mind abilities. Overall, these studies indicate that the development of false belief understanding may correlate with right asymmetric activity, including task-independent resting-state EEG asymmetry and task-dependent functional asymmetry.

It is important to acknowledge that findings from task-independent resting-state EEG asymmetry and task-dependent functional asymmetry should not be equated within the methodologies employed across various studies. The presence of functional activation does not necessarily imply its manifestation in resting state asymmetry patterns. Indeed, previous research has indicated that both forms of brain asymmetry typically exhibit a general pattern of right-lateralized activity associated with the development of Theory of Mind. Notably, some studies, such as Bowman et al. (2019), provided preliminary evidence supporting a potential association between resting-state EEG asymmetry and functional activation asymmetry. However, while Bowman et al. (2019) found preschoolers' resting-state EEG frontal activity at age 4 predicted later ToM-specific fMRI responses in the same brain regions at age 7.5, the applicability of this pattern to children younger than 4 years, and its involvement with right-lateralized asymmetry, remains uncertain.

Previous neural research on Theory of Mind in typically developing individuals has provided evidence linking right hemisphere activity, as measured by resting-state EEG, to the development of Theory of Mind (Sabbagh et al., 2009; Sabbagh & Flynn, 2006). Evidence for the relevance of task-independent asymmetric brain activity to Theory of Mind has also been found in studies involving individuals with autistic spectrum disorder, who show behavioral deficits in Theory of Mind abilities (e.g., Baron-Cohen, 2001). Stroganova et al. (2007) noted atypical broadband resting-state EEG asymmetry in autistic children aged 3 to 8, suggesting a reduced ability of the right temporal cortex for EEG rhythm generation. Brain lesion studies also revealed that damage to the right hemisphere, as opposed to the left, significantly impairs the ability to complete Theory of Mind tasks (Balaban et al., 2016; Griffin et al., 2006; Siegal et al., 1996; Weed et al., 2010; Winner et al., 1998).

In conclusion, from existing neurocognitive research in adults and fewer studies in children, there is indirect evidence for a functional specialization of the right hemispheric brain activity in Theory of Mind reasoning. It is unclear, however, whether this functional specialization is a cause or a consequence of the representational Theory of Mind development (Sabbagh et al., 2009). To address this issue, gathering data on ToM-related brain activation in even younger children, such as infants and toddlers, is necessary. One way of doing so is to use longitudinal designs, where task-independent brain activity is measured in infants or toddlers. Then, children's individual differences in early brain activity can be longitudinally explored for predictive relations with representational Theory of Mind competence when they are around 4 years old. The present exploratory study focuses on one dimension of ToM-related brain activity: right lateralized asymmetry. Our exploratory research questions are whether the task-independent hemispheric asymmetry is associated with Theory of Mind processing during early childhood, whether this asymmetry precedes the development of representational Theory of Mind, and if it can predict later behavioral Theory of Mind performance.

Resting-state EEG provides a reliable measure of task-independent brain activity. Of particular interest has been developmental changes in the ‘‘alpha band’’ (6–9 Hz) of children's resting-state EEG. Alpha waves are thought to relate inversely to brain activity (Gevins, 1998; Klimesch, 1999), implying that higher alpha power in one hemisphere suggests increased activity in the opposite hemisphere (Allen et al., 2004; Reznik & Allen, 2018). Resting-state EEG alpha asymmetry is usually measured by subtracting left from right alpha power, with negative values indicating dominant right hemisphere activity and positive values indicating dominance in the left hemisphere (Allen et al., 2004; Müller et al., 2015, 2018; Sabbagh & Flynn, 2006). Findings on cognitive tasks showed that resting-state EEG alpha asymmetry can predict task performance in a manner consistent with lesion and neuroimaging studies (Hoptman & Davidson, 1998). Roughly 60% of the variance in alpha asymmetry indices is believed to reflect a stable trait-like neural characteristic associated with various psychological constructs (N. A. Fox et al., 1995; Hagemann et al., 2002; Stewart et al., 2011). Research indicates that resting-state EEG frontal alpha asymmetry is a stable characteristic over time across various age groups in children (N. A. Fox et al., 1992; Jones et al., 1997; Müller et al., 2015). For instance, studies reported its stability over a 6–36-month span in children aged 3–6 (Jones et al., 1997; Vuga et al., 2008) and over a 4-year span in ages 4–8 (Kim & Bell, 2006). Furthermore, similar findings were observed over extended periods of up to 69 months across infants and preschoolers. For example, Müller et al. (2015) demonstrated high individual stability of frontal alpha asymmetry from 14 to 83 months of age. Collectively, these findings suggest a consistent pattern of stability in frontal alpha asymmetry across early child development.

False belief understanding is considered the 'litmus test' for evaluating children's Theory of Mind abilities (e.g., Wellman & Woolley, 1990), with significant improvements in explicit false belief understanding tasks observed between ages 3 and 4 (Wellman et al., 2001). The present study investigates the relationship between task-independent asymmetric brain activity before the age of 3 years and behavioral false belief understanding at the age of 4 years. The findings may offer new insights into the neural mechanisms underlying the emergence of Theory of Mind. To our knowledge, research has yet to explore this aspect. The present paper utilizes longitudinal datasets to examine the relationship between asymmetric brain activity and false belief understanding across the developmental trajectory from infancy to 4 years of age. By examining resting-state EEG alpha asymmetry in children at 14 months and 34 months, we aim to determine whether individual differences in relevant asymmetric brain activity, such as those in the frontal and parietal regions, are evident before the emergence of explicit false belief understanding or whether they develop alongside false belief understanding as neural correlates.

Additionally, we aim to better understand the relation between implicit and explicit false belief understanding by including an eye-tracking measure of implicit false belief understanding at 34 and 52 months as an additional outcome variable. A study by Grosse Wiesmann et al. (2020) found that explicit false belief understanding was associated with cortical surface area and thickness in several right-hemispheric regions, including the precuneus, posterior middle temporal gyrus (MTG), and TPJ. In contrast, implicit false belief understanding was linked to a distinct neural network, particularly the right SMG. These findings suggest the possibility of two systems for reasoning about mental states: a more mature, explicit false belief understanding emerging around age 4, and an earlier-developing implicit false belief understanding. Grosse Wiesmann et al. (2020) and other studies (e.g., Bardi et al., 2016; Boccadoro et al., 2019; Hyde et al., 2018) provide preliminary evidence that right-hemispheric regions are involved in Theory of Mind, with distinct areas supporting implicit and explicit false belief understanding. Further research, particularly in younger children, is needed to clarify the developmental trajectories of these neural systems.

The present research used two independent longitudinal datasets, each containing EEG and behavioral assessments of false belief understanding. Study 2 conducted EEG assessments at 34 months and behavioral assessments at 34 and 52 months. It included an anticipatory-looking task to measure implicit false belief understanding at 34 months, and explicit false belief understanding was assessed at 52 months using location and content false belief tasks. Study 2 explores the correlation between resting-state EEG alpha asymmetry and both implicit and explicit false belief understanding. While not initially conceived as a follow-up, Study 3[6] leverages an existing longitudinal

[6] Study 3 analyzed an independent existing dataset to explore the generalizability of findings between resting-state EEG alpha asymmetry and explicit false belief understanding in Study 2;

dataset to complement these explorations. It focuses on earlier resting-state EEG measures at 14 months and their predictive power for later explicit false belief understanding at 51 months. See **Figure 19** for a timeline and sample size of each study. To mitigate concerns about the small sample sizes in each individual study, we also combined the data from both studies to form a larger sample, thereby increasing power and assessing the generalizability of the potential relationship between resting-state EEG alpha asymmetry and explicit false belief understanding. This combined sample represents the largest dataset reported in this research area to date. We acknowledge that the combined sample may introduce concerns such as data independence. Nonetheless, the decision to merge the samples was made to provide a comprehensive analysis given the available data. Moreover, given that children's performance in false belief tasks correlates with their developing language skills (de Villiers & de Villiers, 2014; Milligan et al., 2007) and executive function (Devine & Hughes, 2014), we additionally included standard language skills and executive function assessments.

In summary, our study aims to explore the correlation between resting-state EEG alpha asymmetry and false belief understanding during early childhood. Building on previous studies on the neural correlates of Theory of Mind in young children (e.g., Bowman et al., 2012; Grosse Wiesmann et al., 2020), resting-state EEG brain asymmetry study of Theory of Mind in adults (Sabbagh & Flynn, 2006), and studies involving individuals with autistic spectrum disorder (Stroganova et al., 2007), we pose two exploratory research questions:

---

however, as it was not preregistered, future research should aim to address this limitation. Even so, in the current study, we tried to be as transparent as possible with the procedure for the results of Study 2 and Study 3.

Q1: Is there a correlation between resting-state EEG asymmetry and false belief understanding in children?

Q2: Does the resting-state EEG asymmetry precede the development of representational false belief understanding, and can it predict later behavioral false belief performance?

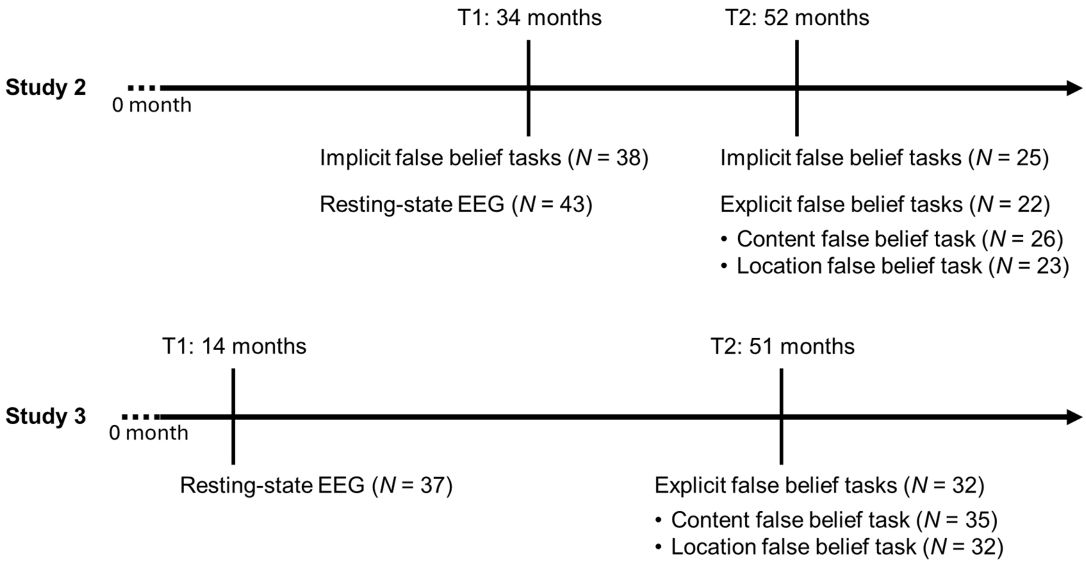


Figure 19. Timeline of the longitudinal study in Study 2 and Study 3. The directional arrow indicates the progression of time from left to right.

3.2. Study 2 methods

3.2.1. Participants

This study was part of a larger longitudinal study[7] for which children were recruited from an urban area in Germany. The final sample included in our research consisted of 43 children (24 girls, M_{age} = 33.79 months, *SD* = 1.25 months, age range = 32.50 - 37.03 months) at Time 1 and 27 children (18 girls, M_{age} = 52.27 months, *SD* = 0.48 months, age range = 51.70-53.83 months) at Time 2. Nine data points from children were missing because children refused to participate in parts of the included tasks (i.e., Implicit false belief task at T1 = 5; Implicit false belief task at T2 = 2; Content false belief task at T2 = 1; Location false belief task at T2 = 1); Three data points from children were missing because of experimenter mistakes (i.e., Location false belief task at T2 = 3). Additionally, 16 children withdrew from T2 data collection. All parents gave written consent after being informed about the experiment procedure. Each child received a personal gift for their participation at each measurement point, and parents were compensated for their travel expenses. The local ethics committee approved the study based on the ethical principles of the European Federation of Psychologists' Associations.

3.2.2. Electrophysiological assessment at T1

3.2.2.1. EEG recording

During brain electrical activity (EEG) recording, children sat quietly on their mother's lap and were presented with brightly colored bubbles on a computer screen (**Figure 20**). This kind of stimulus resembles others used in resting-state EEG research (Kühn-Popp et al., 2016; Licata et al., 2015; Müller et al., 2015; Mundy et al., 2000; Paulus, Kühn-Popp, et al., 2013) and was used to keep the children's visual attention with reduced movement during the

[7] "The role of language in early Theory of Mind development," Crossing the Borders, https://crossing-project.de/, See Kaltefleiter et al., 2021, 2022).

recording time. The recording lasted for at least 3 minutes until the child lost interest in the stimulus, as evidenced by yawning, crying, or strong motor activity.

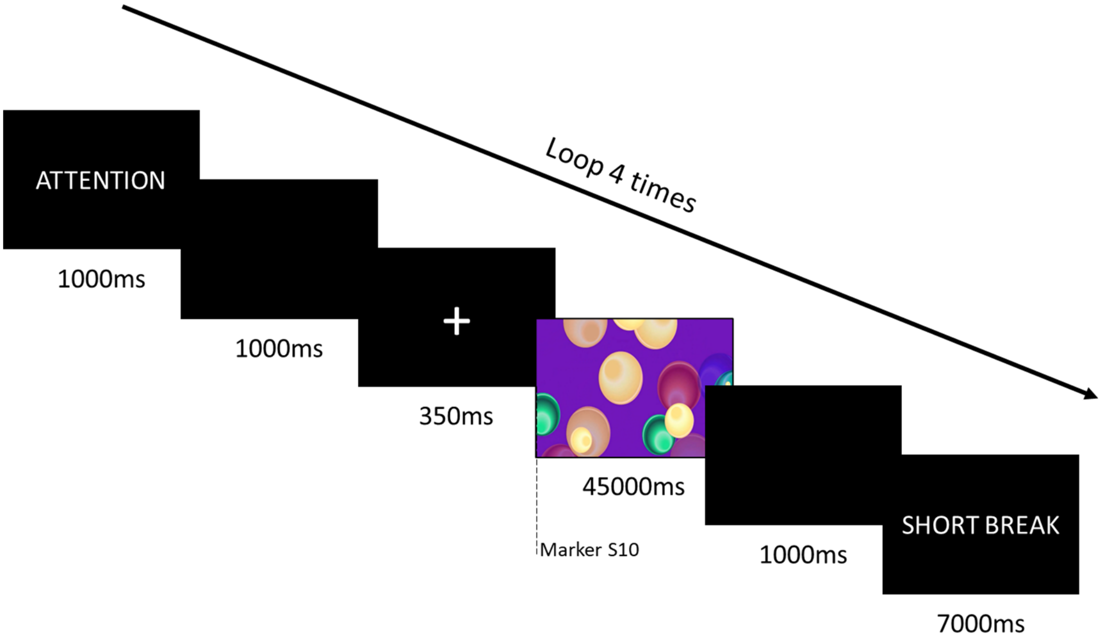


Figure 20. Schematic diagram of resting-state EEG recording.

Recordings were made from 33 electrode sites using an infant-size cap with Ag/AgCl active electrodes (ActiCap, Brain Products, Gilching, Germany) with a layout following the extended international 10-20 system (see **Figure 21** for an overview). The electrical activity from each lead was amplified using the BrainAmp amplifier (Brain Products, Gilching, Germany), sampled at 500 Hz, and referenced to the vertex (Cz). The band passed from 0.016 to 100 Hz, and impedances were kept below 10 kΩ. Fp1 and Fp2 were inserted to detect blinks and vertical eye movements; F9 and F10 were included to detect horizontal eye movements.

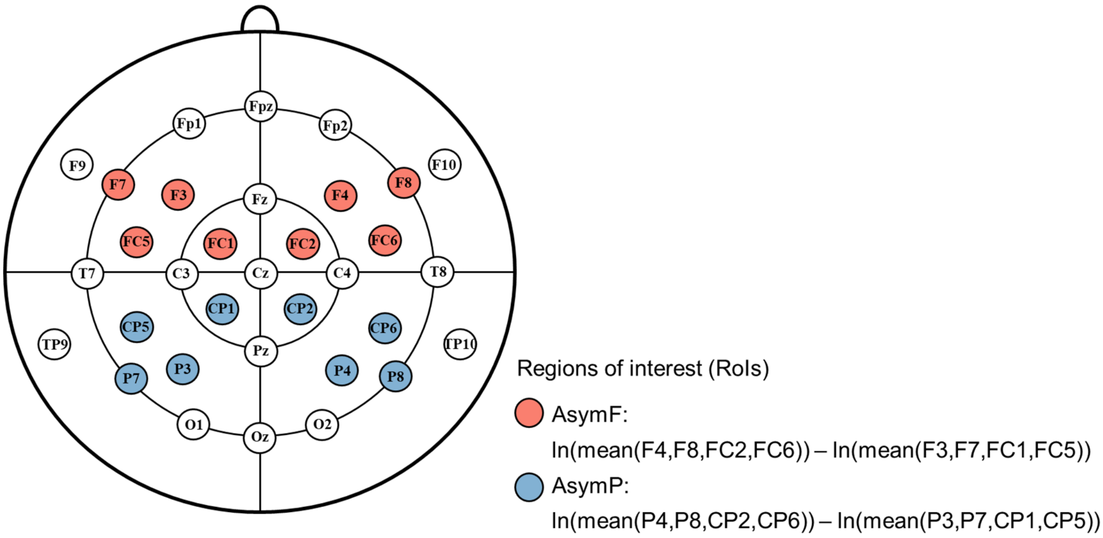


Figure 21. Electrode layout of the resting-state EEG measurement. All channels were included in the EEG measurement. Colored channels show the electrodes included in the data analysis for the resting-state EEG alpha asymmetry scores.

3.2.2.2. EEG analysis

EEG data were analyzed using BrainVision Analyzer Software (Brain Products, Gilching, Germany). Offline, all electrodes were re-referenced to the common average reference, and a digital bandpass filter of 1 to 20 Hz (4^{th} order) was applied. The EEG data were segmented into equal-sized epochs of 1024 ms (512 time points). To ensure that EEG was only analyzed from epochs in which the children were attending to the bubble stimulus, epochs with artifacts were detected through visual inspection and a semiautomatic artifact rejection function (Minimal-maximal difference ± 300μV, amplitude exceeding ± 100μV), and were eliminated from further analyses if they contained eye movements, blinks, or motor artifacts. Data rejection was done blind to conditions by the experimenter. An average of 77.67% of all epochs were included from subsequent analyses, providing on average 133.60 (*SD* = 28.45, range: 61 - 172) epochs per child.

Spectral power, expressed as mean square microvolts (μV^2), was calculated via Fast Fourier Transform (FFT, Hanning window: 10%, frequency resolution: 0.977 Hz). To analyze alpha asymmetries in activation, spectral power (μV^2) was computed for the 6- to 9-Hz (Alpha) frequency band, which corresponds to the frequency band for infants (Kühn-Popp et al., 2016; Licata et al., 2015; Marshall et al., 2002; Müller et al., 2015; Paulus, Hunnius, et al., 2013; Saby & Marshall, 2012). Following previous research investigating EEG alpha asymmetries, EEG power was normalized using natural logarithm transformation (Gasser et al., 1982). Resting-state EEG asymmetry indices were computed by subtracting the ln-transformed EEG power at a given left hemisphere site from the ln-transformed EEG power at its homologous right hemisphere site (Müller et al., 2015, 2018; Sabbagh & Flynn, 2006). That is, the resting-state EEG alpha asymmetry score for the frontal sites (AsymF) was computed by subtracting the average natural logarithm (ln) left power (F3, F7, FC1, FC5) from the average ln right power (F4, F8, FC2, FC6). Similarly, resting-state EEG alpha asymmetry scores for parietal sites (AsymP) were calculated by subtracting left power (P3, P7, CP1, CP5) from right power (P4, P8, CP2, CP6). Electrode sites were chosen based on earlier research investigating children's resting state asymmetries (N. A. Fox et al., 1995; Müller et al., 2015; Paulus, Kühn-Popp, et al., 2013).

3.2.3. Behavioral assessment

From the broader array of tasks integrated into the longitudinal study, only those pertinent to the research aim of the current study were chosen for analysis.

3.2.3.1. Behavioral assessments at Time 1.

Implicit false belief task (Anticipatory-looking false belief task)

The implicit false belief task (Grosse Wiesmann et al., 2017a; Kaltefleiter et al., 2022) was used to assess children's implicit false belief understanding by measuring anticipatory-looking behavior. In the task, children's gaze direction was recorded while watching an agent search for a mouse to assess their implicit tracking of others' beliefs. The task included 10 familiarization trials where the agent consistently found the mouse, and 12 FB trials divided into two different conditions of six trials each. During FB trials, the mouse switched between boxes unknown to the agent, who held a false belief about its location. We concentrated on false belief 1 (FB1) trials in our study[8]. In FB1 trials, the agent observed the mouse's transfer but was unaware of its subsequent departure, continuing to falsely believe the mouse was in its final hiding place despite its actual absence.

Two identical areas of interest (AOIs) were defined for all trials. An overview of the task stimuli and two AOIs can be found in the supplementary materials S1. The "Implicit false belief task DLS score" is a differential looking score (DLS) calculated per trial by subtracting incorrect AOI looking duration from the correct AOI looking duration, divided by the total looking duration at both AOIs (Kaltefleiter et al., 2022). The implicit false belief score is the average of these DLS values across all trials, where a chance performance is 0.

The "Implicit false belief task FF&LL score" combines the average first fixation (FF) and longer look (LL) scores, as established by Grosse Wiesmann et al. (2017). Each trial awards a first fixation score of 1 for looking at the correct AOI first, and 0 for the incorrect one. The longer look score follows similarly: 1 for spending more time at the correct AOI, 0 for the incorrect, and 0.5 for equal

[8] Additionally, our task included another trial type (FB2), wherein the agent did not witness the mouse's transfer and mistakenly believed it remained in the initial location. Our data analysis excluded FB2 trials due to the children's performance in the FB2 trials was significantly below chance. In the FB2 trials, children might not take into consideration that the agent did not watch the target's transfer. Rather, they mostly looked at the last place where they themselves observed the mouse going, neglecting that the agent did not have this information. The detailed methodological reasons can be found in Kaltefleiter et al. (2022).

durations at both. This method of scoring simplifies the assessment process and provides a standardized way to evaluate a child's ability to correctly and rapidly distinguish between AOIs. These discrete values facilitate straightforward comparisons across different trials and participants, ensuring consistency in how responses are evaluated. The overall implicit false belief score per child is computed as the mean of these combined scores, with chance performance for this measure set at 0.5.

3.2.3.2. Behavioral assessments at Time 2.

At T2, two explicit false belief tasks (Wellman & Liu, 2004) were conducted to assess explicit false belief understanding. These tasks were particularly suited for 4-year-olds, aligning with the age when explicit false belief reasoning typically emerges. Furthermore, the identical implicit false belief task administered at T1 was also employed at T2. The explicit false belief sum score at T2 was computed as the sum of the content false belief task score and location false belief task score. The explicit false belief sum score was chosen because the aggregate measures can offer a more reliable and stable assessment of explicit false belief understanding (Grosse Wiesmann et al., 2017a; Kaltefleiter et al., 2022).

Content false belief task

In the content false belief task, children were presented with a Smarties box and asked to speculate about its content. After the children named Smarties as the content, the true content (i.e., a piglet figurine) was revealed. Subsequently, the piglet figurine was returned to the box, and children were asked to identify the content as a memory control. Following this, the figurine Lucas, who had not seen the box's content, was introduced. Children were then posed with a test question (i.e., "What does Lucas think inside the box? Smarties or a piglet?") and a control question (i.e., "Has Lucas looked inside the box before?"). A correct response to both questions earned children one point, resulting in a score range of 0 to 1. The chance level was 25%, as correct responses were required for both the corresponding test and control questions.

Location false belief task

In the location false belief task, children were presented with pictures of a backpack and a closet. They were informed that the figurine Paul was searching for his mittens, which could be in the backpack or the closet. Subsequently, children were informed that Paul's mittens were actually in his backpack, but he falsely believed they were in his closet. The test question asked where Paul would search for his mittens, and the control question inquired about the actual location of Paul's mittens. Children were awarded one point for correctly answering both questions, yielding a score range of 0 to 1. The chance level was 25%, as correct responses were required for both the corresponding test and control questions.

3.2.3.3. Language skills assessment

At T1 and T2, the four age-appropriate subtests of the standardized German language development test for three- to five-year-old children [Sprachentwicklungstest für drei- bis fünfjährige Kinder] (SETK 3-5, Grimm et al., 2015) were administered according to manual instructions. At T1, these subtests encompassed sentence comprehension, encoding semantic relations, phonological working memory, and morphological rule formation. At T2, the subtests included sentence comprehension, sentence memory, phonological working memory, and morphological rule formation. Raw scores obtained in each subtest were converted into standardized T-values using age-specific norm tables. The mean of the T-scores across corresponding subtests (SETK mean score), served as an indicator of children's general language skills.

3.2.3.4. Executive function

A day-night Stroop task (Gerstadt et al., 1994) was conducted at T2 to assess children's inhibition skills as their executive function. Initially, children were required to tell the experimenter when (at night or during the day) the sun/moon and stars are typically in the sky. Subsequently, they were instructed to respond to 'night' when shown a sun card and 'day' for a moon or star card, following a brief practice with two cards. The training phase, involving up to 12 cards, required children to answer correctly on 4 consecutive cards until all cards were presented. Corrective feedback was provided during this phase. In the test phase, no feedback was given, and children's responses to 16 cards were scored based on their correct responses. Incorrect responses, including self-correction and uncertain responses were noted. The total number of correct responses ranged from 0 to 16.

3.2.4. Statistical analysis

Data preparation and all remaining analyses were conducted in IBM SPSS Statistics 29. In light of our exploratory question regarding whether false belief understanding is associated with resting-state EEG alpha asymmetry, two-tailed testing and a significance level of .05 were used for all analyses.

3.3. Study 2 results

3.3.1. Descriptive statistics

Descriptive and inferential statistics are presented in **Table 10** and **Table 11**. Independent sample t tests were conducted to examine potential gender effects on performance in all false belief task scores. No significant gender effects were observed (all *p's* > .05; see supplementary materials S2 for detailed results).

Table 10. Descriptive statistics of the tasks at T1 and one-sample t-tests again chance performance in Study 2.

	N	*M*	*SD*	Value Range	Test Statistic	*p*
Implicit FB task DLS score at T1	38	0.11	0.32	-1 to 1	$t(37) = 2.11$	$p = .042$ [a]
Implicit FB task FF&LL score at T1	38	0.56	0.20	0 to 1	$t(37) = 1.95$	$p = .059$ [a]
SETK3 Comprehension of sentences	39	52.33	8.41	39 - 71		
SETK3 Encoding semantic relations	39	49.08	9.08	32 - 72		
SETK3 Phonological working memory	35	57.83	7.22	44 - 74		
SETK3 Morphological rule formation	37	53.78	8.68	29 - 68		
SETK3 mean score	39	52.76	6.90	36.50 - 66.50		
AsymF [b]	43	-0.02	0.09	-0.26 - 0.20		
AsymP [b]	43	0.00	0.11	-0.17 - 0.30		

Note. FB = false belief. [a] One-sample t-test against chance level (Shapiro-Wilk normality test, *p's* > .05). [b] Negative resting-state EEG alpha asymmetry scores indicate greater relative right than left cortical activity.

Table 11. Descriptive statistics of the tasks at T2 and binomial tests, respectively, one-sample t-tests again chance performance in Study 2.

	N	*M*	*SD*	Value Range	Test Statistic	*p*
Implicit FB task DLS score at T2	25	0.13	0.21	-1 to 1	$t(24) = 3.11$	$p = .005$ [a]
Implicit FB task FF&LL score at T2	25	0.60	0.17	0 to 1	$t(24) = 3.10$	$p = .005$ [a]
Content FB task score at T2	26	0.73	0.45	0 - 1		$p < .001$ [b]
Location FB task score at T2	23	0.65	0.49	0 - 1		$p < .001$ [b]
Explicit FB sum score at T2	22	1.36	0.73	0 - 2		
SETK4 Comprehension of sentences	25	55.92	8.60	40 - 72		
SETK4 Sentence memory	24	54.00	11.36	35 - 81		
SETK4 Phonological working memory	24	54.79	8.05	36 - 69		
SETK4 Morphological rule formation	26	51.58	10.13	31 - 80		
SETK4 mean score	26	53.51	6.42	40 - 63.25		
EF (day-night Stroop task score)	24	10.79	4.73	1 - 16		

Note. FB = false belief. EF = executive function. [a] One-sample t-test against chance level (Shapiro-Wilk normality test, *p's* > .05). [b] Binomial test against chance performance.

3.3.2. Correlation results

To answer the question of whether there are correlations between resting-state EEG alpha asymmetry scores and behavioral false belief task scores, we first examined Pearson correlations between resting-state EEG alpha asymmetry scores at T1 and behavioral false belief task scores at both time points. We also examined Spearman's rank correlations between implicit false belief understanding and explicit false belief understanding. Missing data were pairwise deleted. All detailed results can be found in **Table 12** and **Table 13**.

3.3.2.1. Correlation between resting-state EEG alpha asymmetry scores and implicit and explicit false belief task scores.

To answer the question of whether there are correlations between resting-state EEG alpha asymmetry scores and behavioral false belief task scores, we first examined Pearson correlations between resting-state EEG alpha asymmetry scores at T1 and behavioral false belief task scores at both time points. To further investigate the relationships between resting-state EEG alpha asymmetry scores and implicit as well as explicit false belief task scores, we then calculated partial correlations between resting-state EEG alpha asymmetry scores and behavioral false belief task scores, controlling for co-developing factors (e.g., age, language, executive function). For detailed information on the correlations between co-developing variables, behavioral false belief task scores, and resting-state EEG alpha asymmetry scores, please refer to Supplementary Materials S3. Missing data were pairwise deleted.

Results showed a significant negative correlation between resting-state EEG alpha asymmetry scores of parietal sites and the implicit false belief scores (DLS score: r = -0.382, p = .018, N = 38; FF&LL score: r = -0.531**, p < .001, N = 38) at T1. The greater relative right than left parietal activity, the higher scores were in the implicit false belief task at T1 (see **Table 12**). After controlling for age and language, the partial correlation between parietal resting-state EEG alpha asymmetry and implicit false belief scores is still significant (DLS score: *r-partial* = -0.374, *p* = .025, *N* = 38; FF&LL score: *r-partial* = -0.538, *p* < .001, *N* = 38).

For the correlation between resting-state EEG alpha asymmetry scores and the explicit false belief sum score, results showed a significant negative correlation between resting-state EEG alpha asymmetry scores of frontal sites at T1 and the explicit false belief sum score at T2 (*r* = -0.444, *p* = .039, *N* = 22). The greater the relative right than left frontal activity, the higher the explicit false belief sum score (see **Table 12**). After controlling for age, language, and executive function, the partial correlation between frontal resting-state EEG alpha asymmetry and explicit false belief sum score is still significant (*r-partial* = -0.528, *p* = .036, *N* = 19).

3.3.2.2. Correlation between implicit and explicit false belief task scores.

Finally, we analyzed the correlation between implicit and explicit false belief task scores for both age groups (see **Table 13**). At T2, there was a cross-sectional significant positive correlation between children's implicit false belief task scores and the explicit false belief task sum score (*r's* ≥ 0.438, *p's* ≤ .047, *N* = 21). After controlling for age, language, and executive function, the partial correlation between implicit false belief scores and explicit false belief sum score at T2 is still significant (DLS score: *r-partial* = -0.507, *p* = .038, *N* = 20; FF&LL score: *r-partial* = -0.527, *p* = .030, *N* = 20). However, no longitudinal correlation between children's implicit false belief task scores at T1 and the explicit false belief task sum score was found at T2 (*r's* ≥ -0.188, *p's* ≥ .401, *N* = 22). These results were consistent with the findings reported by Kaltefleiter et al. (2022), indicating a cross-sectional but not a longitudinal correlation between implicit false belief understanding and explicit false belief task sum score.

Table 12. Correlations between resting-state EEG alpha asymmetry scores and false belief task scores in Study 2.

	AsymF	AsymP
Implicit FB task DLS score at T1 (*N* = 38)	0.159, *p* = .340	**-0.382*, *p* = .018**
Implicit FB task FF&LL score at T1 (*N* = 38)	0.160, *p* = .338	**-0.531**, *p* < .001**
Implicit FB task DLS score at T2 (*N* = 25)	-0.311, *p* = .130	-0.070, *p* = .740
Implicit FB task FF&LL score at T2 (*N* = 25)	-0.175, *p* = .402	-0.243, *p* = .241
Explicit FB sum score at T2 (*N* = 22)	**-0.444*, *p* = .039**	-0.016, *p* = .944

Note: † *p* < .10; * *p* < .05; ** *p* < .01; *** *p* < .001 (two-tailed). FB = false belief. EF = executive function. Sample sizes are presented in parentheses.

Table 13. Correlations between implicit and explicit false belief understanding in Study 2.

	Implicit FB DLS score at T1	Implicit FB FF&LL score at T1	Implicit FB DLS score at T2	Implicit FB FF&LL score at T2
Explicit FB sum score at T2	-0.290, *p* = .190 (*N* = 22)	-0.188, *p* = .401 (*N* = 22)	**0.438*, *p* = .047 (*N* = 21)**	**0.534*, *p* = .013 (*N* = 21)**

Note: † *p* < .10; * *p* < .05; ** *p* < .01; *** *p* < .001 (two-tailed). FB = false belief. EF = executive function. Sample sizes are presented in parentheses.

3.4. Interim discussion on Study 2

In Study 2, we conducted a longitudinal exploration of the relationships between resting-state EEG alpha asymmetry and the development of Theory of Mind abilities, specifically implicit and explicit false belief understanding, in children aged 3 to 4 years. Our study employed a longitudinal design to assess resting-state EEG alpha asymmetry at 34 months (T1) and examine its relationship with explicit false belief task performance at 52 months (T2), thereby tracing the predictive value of early neural activity patterns for later false belief understanding. Implicit false belief tasks were assessed at both time points. Results indicate that resting-state EEG alpha asymmetry correlated both with implicit and explicit false belief understanding: Firstly, greater relative right than left parietal activity was associated with better performance in an implicit false belief task at T1. Secondly, greater relative right than left frontal activity was significantly correlated with enhanced performance in explicit false belief tasks at T2. Importantly, these correlations (i.e., the correlation between implicit and explicit false belief understanding and resting-state EEG alpha asymmetry) were noted independently of age and other cognitive abilities commonly associated with false belief understanding, including language skills and executive function.

These results showed correlations between false belief understanding and resting-state EEG alpha asymmetry, indicating that there may be hemispheric asymmetry in false belief understanding processing. Furthermore, such asymmetry may manifest early in development, even before the emergence of false belief understanding. The distinctions we observed between the neurodevelopmental trajectories of implicit and explicit false belief understanding delineate a nuanced relationship between resting-state EEG alpha asymmetry and false belief understanding during early childhood. Nevertheless, given the relatively small sample size in Study 2, these findings should be regarded as preliminary evidence pointing potentially to early-appearing neural precursor underlying false belief understanding between ages 3 and 4. Further research with an independent dataset will be crucial to substantiate and extend initial findings.

3.5. Study 3 and combined samples

To further support the preliminary evidence from Study 2 concerning explicit false belief understanding in a limited cohort, another independent longitudinal dataset [9] was included. We then analyzed Study 3 and the combined samples from Study 2 and Study 3 to replicate and strengthen the initial findings on the relationship between resting-state EEG alpha asymmetry and explicit false belief understanding.

3.6. Study 3 methods

In Study 3, resting-state EEG data collection took place with 14-month-old infants. Furthermore, the age at which the explicit false belief tasks (i.e., a content false belief task and a location false belief task) were administered aligned with that in Study 2, that is, around the age of 51 months. No implicit false belief understanding was assessed in this study. In both Study 2 and Study 3, language skills and executive function measures were employed.

3.6.1. Participants

The final sample for Study 3 comprised 37 children (18 girls, M_{age} = 13.99 months, SD = 0.22 months, age range = 13.60 – 14.43 months) at Time 1 and 35 children (17 girls, M_{age} = 50.72 months, SD = 0.99 months, age range = 49.70 – 54.57 months) at Time 2. An additional three children participated in the resting-state EEG recording but were excluded from the final sample due to fussiness. These children were drawn from a larger cohort in a longitudinal study, Theory of Mind in Infancy and Early Childhood (TOMII/TOMEC) study,

[9] Longitudinal Study on Theory of Mind in Infancy and Early Childhood, see Sodian et al. (2020).

which assessed Theory of Mind across multiple time points (see Kloo et al., 2022; Kloo & Sodian, 2017; Osterhaus et al., 2022, for a further description of the sample). Parental written consent was obtained before participation. The local ethics committee approved the study based on the ethical principles of the European Federation of Psychologists' Associations.

3.6.2. Electrophysiological assessment at T1

3.6.2.1. EEG recording

The EEG recording setup for the resting state in Study 3 paralleled that of Study 2, adjusted for a younger cohort with an EEG cap sized for 14-month-old infants. Data were collected from 17 electrode sites (Fp1, Fp2, F3, F4, F7, F8, F9, F10, C3, Cz, C4, T7, T8, P3, P4, O1, O2) using an infant-specific cap with Ag/AgCl active electrodes (ActiCap, Brain Products, Gilching, Germany), following the 10/20 system guidelines.

3.6.2.2. EEG analysis

EEG data were examined and analyzed using BrainVision Analyzer Software (Brain Products, Gilching, Germany). Offline, all electrodes were re-referenced to the common average reference, and a digital bandpass filter of 1 to 20 Hz (4^{th} order) Hz was applied. The EEG data were segmented into equal-sized epochs of 1024 ms (512 time points). To ensure that EEG was only analyzed from epochs in which the children were attending to the bubble stimulus, epochs with artifacts were detected through visual inspection and a semiautomatic artifact rejection function (Minimal-maximal difference ± 300 µV, amplitude exceeding ± 120 µV), and were eliminated from further analyses if they contained eye movements, blinks, or motor artifacts. Data rejection was done blind to conditions by the experimenter. On average, 56.39 % of all epochs were included from subsequent analyses, providing on average 149.51 (*SD* = 48.07, value range = 68 - 249) epochs per infant.

Artifact-free epochs were extracted through a Hanning window and power spectra were calculated using the Fast Fourier Transform. To analyze asymmetries in activation, power (μV^2) was computed for the 6 – 9 Hz (alpha) frequency band. A grand average of the FFTs was calculated for every participant. Subsequently, an asymmetry score for frontal sites (AsymF) was calculated by subtracting the average natural logarithm (ln) left power (F3, F7) from the average ln right power (F4, F8). Similarly, asymmetry scores for parietal sites (AsymP) were calculated by subtracting left power (P3) from right power (P4).

3.6.3. Behavioral assessment

The assessed explicit false belief tasks (i.e., a content false belief task and a location false belief task) and the language skill assessment (i.e., SETK4) were the same as in Study 2.

We used the 'Simon Says' task to measure executive function (Strommen, 1973). Children were engaged in a game with clear instructions: "Now, we are playing a game. I'll do all the exercises. Sometimes you are to do them with me and sometimes you are not. Only if I say, 'Simon says,' you do them. If I don't, you don't do them." Children initially practiced through one Simon and one non-Simon trial with corrective feedback to ensure understanding. This was followed by 20 test trials, evenly split between Simon and non-Simon types, without feedback. The experimenter executed all actions (e.g., 'touch your nose', 'stamp your feet') in a predetermined yet randomized order, with a rule reminder after the first 10 trials. Performance was rated on a 0-2 scale per trial: a '2' for correctly following or ignoring commands based on trial type, a '1' for partial or incorrect responses, and a '0' for failing to adhere to instructions. Scores were summed for each trial type, allowing a maximum of 20 points per type. Only non-Simon trial scores were evaluated for analysis.

3.6.4. Statistical analysis

All data analyses were the same as in Study 2.

3.7. Results in Study 3 and combined samples

3.7.1. Descriptive statistics in Study 3

Descriptive performance on all tasks is presented in **Table 14**. Independent sample t-tests were conducted to examine potential gender effects on performance in all false belief task scores. No significant gender effect was observed ($p > .05$; see the supplementary materials S2 for detailed results).

Table 14. Descriptive statistics of resting-state EEG alpha asymmetry and behavioral tasks performance in Study 3.

	N	*M*	*SD*	Value Range	*p*
Content FB task at T2	35	0.31	0.47	0 - 1	*p* = .242 [a]
Location FB task at T2	32	0.41	0.50	0 - 1	*p* = .038 [a]
Explicit FB sum score at T2	32	0.72	0.73	0 - 2	
SETK mean score	34	55.88	10.55	32 - 74	
EF (Simon Says task score)	31	6.16	6.25	0 - 18	
AsymF	37	-0.07	0.24	-0.78 – 0.40	
AsymP	37	-0.02	0.47	-1.97 – 0.88	

Note: FB = false belief. EF = executive function. [a] Binomial test against chance performance.

3.7.2. Correlations between resting-state EEG alpha asymmetry scores and explicit false belief sum scores in Study 3

For the correlation between resting-state EEG alpha asymmetry scores and explicit false belief sum score (see **Table 15**), results showed a significant negative correlation between resting-state EEG alpha asymmetry scores of frontal sites at T1 and explicit false belief sum score at T2 ($r = -0.450$, $p = .010$, $N = 32$). The greater the relative right than left frontal activity, the higher the explicit false belief sum score. The partial correlation analyses, controlling for age, co-developing language skills, and executive function simultaneously, yielded a marginal significance (*r-partial* = -0.344, *p* = .092, *N* = 28).

Table 15. Longitudinal correlations between resting-state EEG alpha asymmetry scores at 14 months and explicit false belief sum score at 51 months in Study 3.

	AsymF	AsymP
Explicit FB sum score at T2 (*N* = 32)	**-0.450**, *p* = .010**	-0.190, *p* = .298

Note: FB = false belief. EF = executive function. * *p* < .05; ** *p* < .01; *** *p* < .001 (two-tailed). In bold font: *p* < .05 FDR corrected.

3.7.3. Results on explicit false belief understanding from combined data of Study 2 and Study 3

Study 3 is an independent longitudinal dataset regarding the explicit false belief understanding and resting-state EEG alpha asymmetry. To consolidate these findings in Study 2 regarding the explicit false belief understanding, we further combined the samples from Study 2 and Study 3, resulting in a total of *N* = 80 children (42 girls, M_{age} = 24.63 months, *SD* = 9.97 months, age range = 13.60 – 37.03 months) who completed the resting-state EEG recording, and 54 children (29 girls, M_{age} = 51.34 months, *SD* = 1.14 months, age range = 49.70 – 54.57 months) who completed the explicit false belief tasks. Analysis of this combined sample revealed a significant negative correlation between resting-state EEG alpha asymmetry scores of frontal sites and the explicit false belief sum score, indicating that greater relative right than left frontal activity correlates with a higher explicit false belief sum score (see **Table 16**. *r* = -0.282, *p* = .039, *N* = 54). After controlling for age, language, and executive function[10], the partial correlation between frontal resting-state EEG alpha asymmetry and explicit false belief sum score still persisted (*r-partial* = -0.327, p = .031, N = 47).

Table 16. Longitudinal correlations between resting-state EEG alpha asymmetry scores and explicit false belief performance in combined samples.

	AsymF	AsymP
Explicit FB sum score (N = 54)	**-0.282*, *p* = .039**	-0.109, p = .434

Note: FB = false belief. EF = executive function. * p < .05; ** p < .01; *** p < .001 (two-tailed).

To further explore the relationship between resting-state EEG alpha asymmetry and explicit false belief understanding, hierarchical linear regression models [11] were conducted to examine whether AsymF explained additional variance in explicit false belief understanding after accounting for age, language skills, and executive function (see **Table 17**). Model 1, which included age, language skills, and executive function, was significant, $F(3,43) = 5.959$, p = .002, and explained approximately 24.4% of the variance in explicit FBU (*Adjusted* R^2 = .244). Model 2, which added AsymF to the predictors, was also significant, $F(4,42) = 6.139$, p < .001, and explained 30.9% of the variance in explicit false belief understanding (*Adjusted* R^2 = .309). The inclusion of AsymF in Model 2 significantly improved the explained variance in explicit false belief

[10] Given the differing assessments of executive function in Study 2 and Study 3, we utilized the Z-score of executive function for the combined sample analyses.

[11] A commonly cited rule of thumb for sample size in multiple regressions suggests a minimum of 10 observations per predictor variable (Howell, 2010). With 4 predictor variables, this would require a minimum sample size of 40. Therefore, the regression model, incorporating age, language skills, executive function, and resting-state EEG alpha asymmetry, was conducted using the combined samples (N = 47). Due to the limited sample sizes of Study 2 and Study 3 individually, separate regression analyses were not included in the current manuscript.

understanding (ΔR^2 = .075, $\Delta F(1, 42)$ = 5.012, p = .031). This suggests that frontal asymmetry contributes unique variance to the prediction of explicit false belief understanding beyond age, language skills, and executive function.

These findings from the combined dataset are consistent with those of Study 2 and Study 3, strengthening the evidence that superior explicit false belief understanding is associated with greater right than left frontal activity. Although these results are promising, they should be viewed as supplementary, as using different EEG cap sizes for the 34-month-old and 14-month-old children introduced challenges in defining consistent resting-state EEG alpha asymmetry across age groups. Despite these limitations, the overall pattern of results aligns with our initial hypotheses.

Table 17. Results of hierarchical linear regression predicting the explicit false belief sum score at T2 from resting-state EEG alpha asymmetry score at T1 in combined samples.

Model	Variable	Estimate	*SE*	*Beta*	*t*	*p*	*Adjusted* R^2
Model1	(Intercept)	-1.007	0.708		-1.423	.162	$.244^{**}$
	Age	0.045	0.011	0.524	4.047	< .001	
	Language	0.018	0.011	0.198	1.535	.132	
	EF	0.098	0.104	0.121	0.940	.353	
Model2	(Intercept)	-0.744	0.687		-1.083	.285	$.309^{***}$
	Age	0.049	0.011	0.568	4.535	< .001	
	Language	0.010	0.011	0.116	0.901	.373	
	EF	0.067	0.100	0.082	0.663	.511	
	AsymF	-1.290	0.576	-0.294	-2.239	.031	

Note: * p < .05; ** p < .01; *** p < .001. EF = executive function.

3.8. Discussion of results on explicit false belief understanding from Study 3 and the combined samples

Results analyses from Study 3 and the combined data confirmed that superior explicit false belief performance at age 4 years was associated with greater right than left frontal activity at a younger age. However, the partial correlation in Study 3 decreased to marginal significance when controlling for age, language, and executive function simultaneously. We attribute this reduction to the limited sample size, which may affect statistical power. This interpretation was supported by the results from the combined sample, where right frontal activity remained significantly related to representational false belief performance when age, language, and executive function were controlled for simultaneously. Nevertheless, to support this assumption, a replication with a larger sample is required. These findings highlight the potential influence of individual differences in task-independent frontal activity on the development of explicit false belief understanding; however, it is important to interpret these findings with caution due to the limited sample size. In the following Discussion, we will synthesize these results to discuss the relationship between resting-state EEG alpha asymmetry and false belief understanding development more comprehensively.

3.9. Discussion on Studies 2 and 3

Around the age of 4 years, children begin to develop the representational Theory of Mind, enabling them to understand others' false beliefs and the influence of these beliefs on actions. The present study explored the neural mechanisms underlying the representational Theory of Mind during early childhood and offers preliminary evidence of longitudinal continuity in the brain systems that support Theory of Mind reasoning. Specifically, through two independent longitudinal studies, this study explored the relationship between resting-state EEG alpha asymmetry and behavioral false belief competence during early childhood. Results identified a pattern in which greater right frontal activity at 14 and 34 months is longitudinally associated with better performance on explicit false belief tasks at age 4. This pattern tentatively suggests that the brain lateralization in resting-state EEG alpha power emerges before the observable behavioral manifestations of Theory of Mind. Additionally, superior implicit false belief understanding correlated with greater right parietal activity, which could preliminarily suggest that the neural mechanisms underlying implicit and explicit false belief understanding may be partly distinct. For a visual summary of the results please see **Figure 22**.

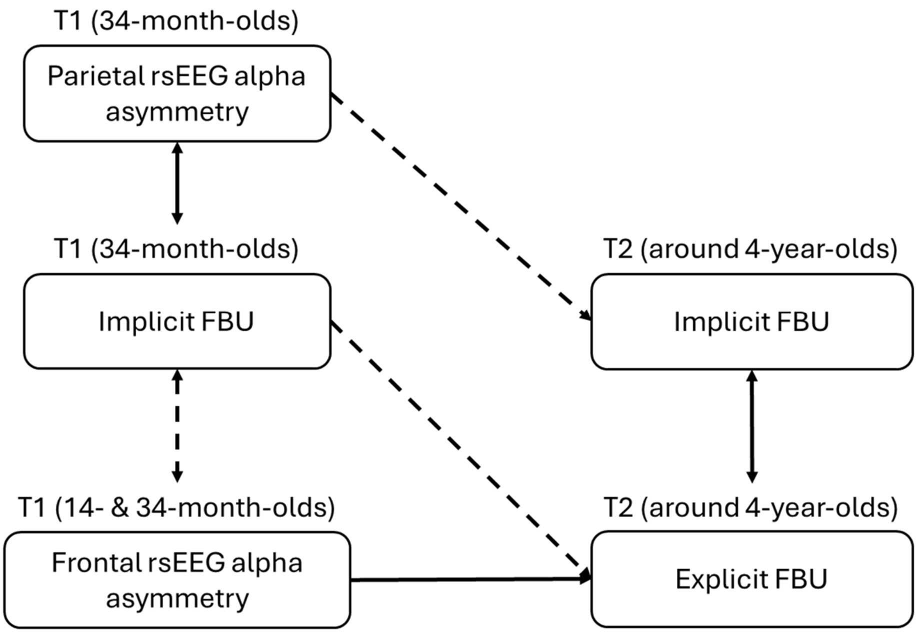


Figure 22. Visual summary of the correlations between T1 and T2 in Study 2 and Study 3. Solid lines indicate statistically significant correlations, while dotted lines denote correlations that did not reach statistical significance.

It is important to highlight that the present findings are preliminary and exploratory due to limitations in statistical power because of the low sample size. Although we augmented our sample size by combining data from two studies, future research is necessary to include larger samples to replicate and confirm these findings. Nevertheless, given the limited focus of previous research on neural activity during the sensitive periods of false belief understanding development in infancy/toddlerhood, we think our longitudinal study provides valuable insights into the potential relationship between resting-state EEG alpha asymmetry and false belief understanding development. In the following sections, we will discuss and cautiously interpret the key findings of this research and offer suggestions for future research. All our interpretations should be read in light of the small sample size.

The finding that greater right than left frontal activity, observable in infancy and toddlerhood, predicts explicit false belief understanding competence by age 4 tentatively supports the view that the lateralization involved in the functional specialization of Theory of Mind reasoning in the brain is a neural marker that appears prior to the behavioral development of Theory of Mind. Our results suggest longitudinal continuity and align with previous, albeit limited, cross-sectional studies that identified the right MPFC as associated with the representational Theory of Mind reasoning in children. Notably, Bowman et al. (2012) observed right mid-frontal activations linked to mental-state reasoning in 7- to 8-year-old children, suggesting a shared neural pattern across tasks involving reasoning about beliefs and desires. While the functional asymmetry of Theory of Mind activity may not directly correspond to its anatomical asymmetry, developmental neural research indicates that false belief competence correlates with anatomical changes, including increased fractional anisotropy in the white matter of the right MPFC around ages 3 to 4 (Grosse Wiesmann et al., 2017b). Our data extended these prior findings and carefully suggest that task-independent right frontal brain activity in children younger than 3 years old is associated with better Theory of Mind behavioral responses at age 4.

Furthermore, our longitudinal data showed a stable correlation between early-appearing right frontal activity and later explicit false belief understanding across early childhood. Previous research using the resting-state EEG technique has demonstrated high individual stability of frontal alpha asymmetry from 14 to 83 months of age (Müller et al., 2015). Based on these findings, we propose that neural stability in frontal asymmetry is likely to persist from 14 to 34 months. Our results support this proposition, suggesting that explicit false belief understanding at age 4 exhibits significant longitudinal correlations with frontal resting-state EEG alpha asymmetry, regardless of whether it is assessed at 14 months or 34 months. Intriguingly, the stability on the neural level occurs despite notable advancements in explicit false belief understanding, as demonstrated by behavioral tests where 4-year-olds performed above chance, even though they had not developed explicit false belief understanding 1 or 3 years earlier. The neural consistency observed in our data, amid changes in conceptual performance, suggests that the cognitive capacities supported by frontal regions are probably established earlier in development and remain stable into early childhood. Importantly, Bowman et al. (2019) found that frontal resting-state EEG brain activity by age 4 correlates with specialized neural responses in the same brain regions at 7.5 years. This finding provided preliminary evidence that the early developments in the region of dorsal MPFC that are important for Theory of Mind in 4-year-olds are also associated with the extent of functional specialization of that same region for Theory of Mind reasoning 3.5 years later, suggesting early stability in the neural system for Theory of Mind reasoning, and longitudinal continuity in this neural system despite behavioral-cognitive advancements (Bowman et al., 2019). Our study extends this implication for neural longitudinal continuity of explicit false belief understanding prior to age 4, although these results remain tentative and call for further investigation to confirm these early findings.

One key aspect of our study design at both T1 and T2 is that the continuities in frontal resting-state EEG alpha asymmetry relate to Theory of

Mind reasoning. We controlled for age, language skills, and executive function in our design, acknowledging that these factors are known to co-develop and can potentially influence children's performance on explicit false belief tasks. Our results demonstrated that right frontal activity was significantly associated with representational false belief performance, even when these co-developing factors were taken into account. This suggests that the observed longitudinal correlation between frontal resting-state EEG alpha asymmetry and Theory of Mind performance is unlikely to be merely a byproduct of these co-developing constructs. However, to further validate these results, we recommend that future studies incorporate larger sample sizes to ensure adequate statistical power and thereby strengthen the credibility of the evidence.

To briefly summarize, our findings regarding the potential relationship between frontal asymmetric activity and explicit false belief understanding extend prior research on the role of the frontal cortex in explicit false belief understanding. The observed longitudinal association between increased right frontal activity and enhanced explicit false belief understanding provides preliminary insights into the developmental onset and progression of explicit false belief understanding in young children. Nevertheless, considerable caution is warranted when interpreting these results due to the small sample size and the potential for spurious correlations. It is imperative to conduct further research with larger and more generalizable samples, as well as additional evaluation time points along the developmental trajectory, to confirm and expand upon these findings. Additionally, further investigation is required to elucidate the role of right frontal activity in the explicit false belief understanding during early childhood. This inquiry could extend to examining the relationship between right frontal activity and a wider variety of early-emerging socio-cognitive skills, such as perspective-taking (Tullett et al., 2012), self-other distinctions [12] (Schuwerk et al., 2014), and cognitive empathy (Rueckert & Naybar, 2008), which previous research has linked to Theory of Mind. It is pertinent to explore whether right frontal activity might underpin a broader spectrum of ToM-related functions from an early age (Krause et al., 2012; Schuwerk et al., 2014; Sebastian et al., 2012).

Additionally, our study investigated the relationship between task-independent brain activity and implicit false belief understanding (assessed with an anticipatory looking task; see Grosse Wiesmann et al., 2017a; Kaltefleiter et al., 2022). We found a significant correlation between superior performance on the implicit false belief task at the age of 3 years and greater right than left parietal-related activity. This finding is in line with earlier research showing sensitivity to others' beliefs in the parietal regions in 7-month-olds (Hyde et al., 2018). Further emphasizing the importance of the right parietal region in Theory of Mind, Grosse Wiesmann et al. (2020) identified a correlation between implicit false belief competence and increased cortical thickness in the right SMG—a critical part within the parietal regions—among 3- to 4-year-olds. While the spatial resolution of EEG limits direct structural comparisons with brain imaging studies, our results align with the previous research and support the involvement of right parietal activity in the early age range of implicit false belief understanding development.

Findings by Grosse Wiesmann et al. (2017b) and by Sabbagh et al. (2009) also observed right hemispheric activity in parietal regions like the TPJ and precuneus gyrus during resting-state EEG recordings linked to explicit Theory of Mind reasoning in 4-year-olds. These findings are partially inconsistent with the present findings, which suggested a dissociation between

[12] The definition of self-other distinction is "the process by which one distinguishes between self- and other-related representations (cognitive, affective, sensorimotor, etc.)" (Quesque et al., 2024).

explicit and implicit false belief reasoning, showing that frontal alpha asymmetry was associated with explicit false belief understanding while parietal alpha asymmetry was associated with implicit false belief understanding. This finding might be seen as tentative support for the view that two distinct neural systems support implicit and explicit false belief understanding development. However, previous research has emphasized the correlation of right parietal activity in both explicit false belief understanding (e.g., Sabbagh et al., 2009; Saxe & Wexler, 2005) and implicit false belief understanding (e.g., Boccadoro et al., 2019; Hyde et al., 2018), though these processes were associated with distinct subregions of the parietal cortex (Grosse Wiesmann et al., 2020). In our study, greater right parietal activity was significantly correlated with better performance in implicit false belief tasks but not explicit false belief tasks. Caution is warranted, however, in interpreting nonsignificant correlations between parietal resting-state EEG alpha asymmetry scores and explicit false belief understanding. It is important to note that a relatively small sample size may limit the statistical power to detect a significant effect if one exists. Therefore, it is impossible to draw a definitive conclusion about the existence of a relationship based on nonsignificant results, and we suggest future research with a larger sample to further investigate this question. Additionally, the absence of significant longitudinal correlations between parietal asymmetry and explicit false belief understanding at age 4 may also reflect the potential instability of parietal asymmetry (Bowman et al., 2019; Müller et al., 2015). Future investigations should also clarify the stability and influence of parietal asymmetry in implicit and explicit false belief understanding across a broader age range.

While our findings underscore the correlations between resting-state EEG alpha asymmetry and behavioral false belief competencies, they also point to several unresolved questions. It remains unclear to what extent the distinct neural correlates are independent of variations in task formats, particularly between anticipatory looking in implicit false belief tasks and verbally elicited responses in explicit false belief tasks. Furthermore, the precise nature of the observed asymmetrical cortical activity—greater relative right than left activity in parietal regions for implicit false belief tasks and that in frontal regions for explicit false belief tasks—warrants further exploration. Future research should, therefore, directly contrast brain activation during the nuanced processes of the implicit and explicit false belief understanding to determine how neural correlates vary independently of task formats in early childhood.

In summary, drawing on previous neural research, which highlighted asymmetry in functional activation patterns related to Theory of Mind reasoning, our study represents the first exploration into the potential connection between resting-state asymmetric brain activity and false belief understanding during early childhood. Across two independent longitudinal studies and combined samples, we observed preliminary evidence that a greater relative right than left frontal activity at 34 months and even at 14 months was associated with better performance in explicit false belief tasks at 4 years of age. Additionally, enhanced performance in implicit false belief tasks at age 3 appeared to correlate with greater relative right versus left parietal activity. While these findings provide preliminary evidence of potential relationships between resting-state EEG alpha asymmetry in the frontal and parietal regions and false belief understanding, limitations in power and sample size warrant that these results be interpreted as exploratory. We attempted to alleviate concerns related to limited sample size by combining samples; however, further research is needed to confirm and refine these findings. Nonetheless, given the rarity of resting-state EEG data collected from infants and toddlers, these results may serve as a valuable initial exploration for future research in this field.

Supplementary materials

S1. An overview of the task stimuli and two AOIs in the implicit false belief task of Study 2.

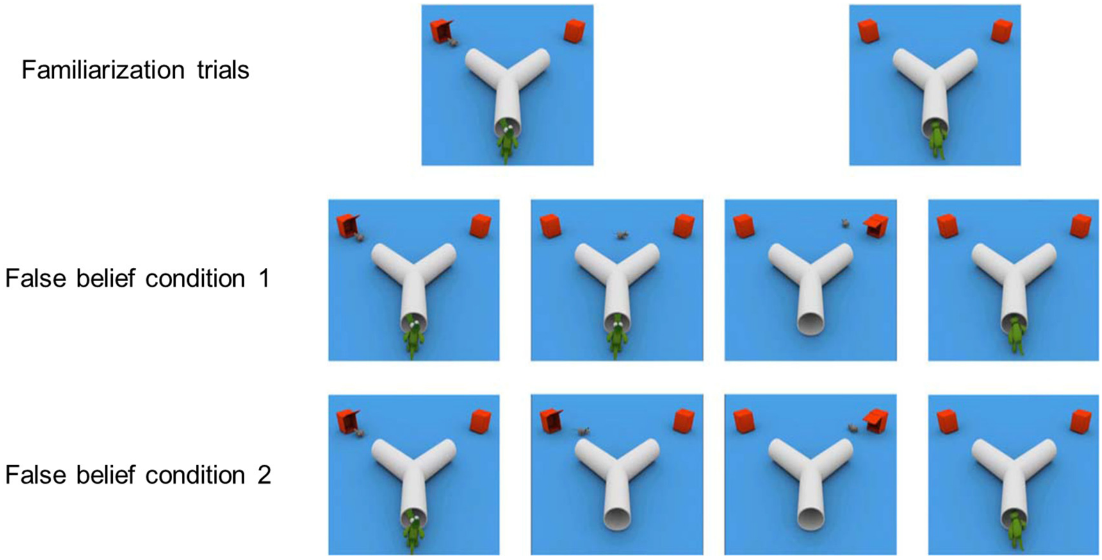


Figure S1. Schematic diagram of implicit false belief task (Anticipatory-looking false belief task) (Kaltefleiter et al., 2022).

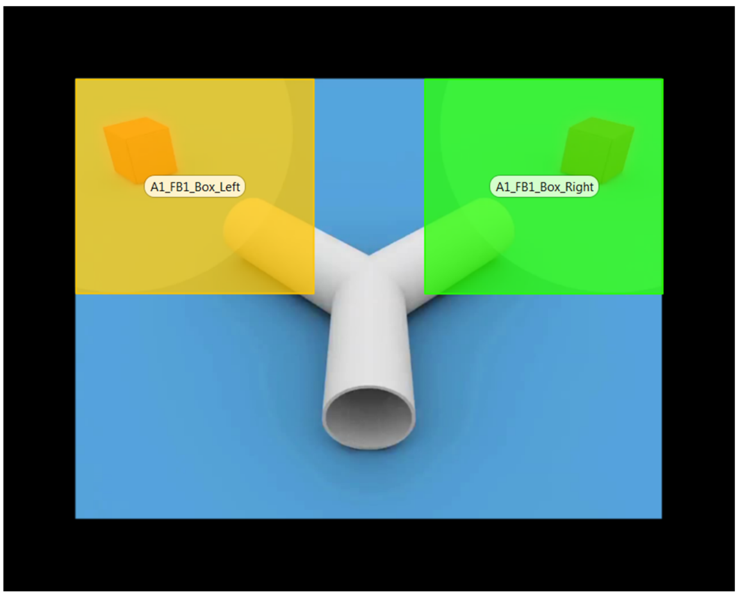


Figure S2. Region of Interest (ROI) in the implicit false belief task (Kaltefleiter et al., 2022).

S2. Independent sample t tests were conducted to examine potential gender effects on performance in all false belief task scores.

Table S1. Independent sample t tests for gender effect in Study 2.

Dependent Variables	*t*	*df*	*p (two-tailed)*
1. Implicit FB task DLS score at T1	-0.19	36	.850
2. Implicit FB task FF&LL score at T1	-0.70	36	.491
3. Implicit FB task DLS score at T2	-1.22	23	.235
4. Implicit FB task FF&LL score at T2	-1.67	23	.109
5. Explicit FB sum score at T2	-0.55	20	.592

Note: FB = false belief.

Table S2. Independent sample t tests for gender effect in Study 3.

Dependent Variables	*t*	*df*	*p (two-tailed)*
Explicit FB sum score at T2	0.59	30	.562

Note: FB = false belief.

Table S3. Independent sample t tests for gender effect in combined samples.

Dependent Variables	*t*	*df*	*p (two-tailed)*
Explicit FB sum score at T2	-1.23	52	.224

Note: FB = false belief.

S3. The correlations between co-developing variables, behavioral false belief task scores, and resting-state EEG alpha asymmetry scores

Table S4. The correlation between co-developing variables (T1), behavioral false belief task scores (T1), and resting-state EEG alpha asymmetry scores in Study 2.

	Age	Language
Implicit FB task DLS score at T1	0.17, *p* = .305	0.03, *p* = .880
Implicit FB task FF&LL score at T1	0.04, *p* = .835	0.00, *p* = .984
AsymF	-0.04, *p* = .783	0.12, *p* = .455
AsymP	-0.05, *p* = .777	0.17, *p* = .290

Table S5. The correlation between co-developing variables (T2), behavioral false belief task scores (T2), and resting-state EEG alpha asymmetry scores in Study 2.

	Age	Language	Executive function
Implicit FB DLS score at T2	-0.10, *p* = .632	-0.25, *p* = .235	0.26, *p* = .226
Implicit FB FF&LL score at T2	0.12, *p* = .579	-0.37, *p* = .079	0.33, *p* = .124
Explicit FB sum score at T2	0.40, *p* = .060	-0.16, *p* = .477	0.32, *p* = .163
AsymF	0.11, *p* = .574	0.12, *p* = .572	-0.30, *p* = .159
AsymP	-0.05, *p* = .814	0.23, *p* = .265	-0.31, *p* = .135

Note: * $p < .05$; ** $p < .01$; *** $p < .001$ (two-tailed).

Table S6. The correlation between co-developing variables, explicit false belief sum score, and resting-state EEG alpha asymmetry scores in Study 3.

	Age	Language	Executive function
Explicit FB sum score at T2	-0.24, *p* = .192	0.44*, *p* = .011	-0.04, *p* = .828
AsymF	0.28, *p* = .109	-0.35*, *p* = .042	-0.11, *p* = .541
AsymP	0.03, *p* = .858	-0.01, *p* = .977	-0.04, *p* = .824

Note: * *p* < .05; ** *p* < .01; *** *p* < .001 (two-tailed).

Table S7. The correlation between co-developing variables, explicit false belief sum score, and resting-state EEG alpha asymmetry scores in combined samples.

	Age	Language	Executive function
Explicit FB sum score	0.40**, *p* = .003	0.18, *p* = .193	0.08, *p* = .577
AsymF	0.15, *p* = .191	-0.29*, *p* = .022	-0.15, *p* = .272
AsymP	0.04, *p* = .756	0.01, *p* = .960	-0.09, *p* = .516

Note: * *p* < .05; ** *p* < .01; *** *p* < .001 (two-tailed).

Chapter 4: General Discussion

The ability to attribute mental states to others, known as Theory of Mind, is fundamental to human social cognition. A quintessential measure of Theory of Mind ability is understanding others' false beliefs about the world. This dissertation comprises three empirical studies that collectively aim to make novel contributions to identifying the neural correlates of false belief understanding in children under the age of 4 years, an under-researched area. These studies are the first to investigate task-dependent ERP correlates of false belief understanding in toddlers and the first to explore the task-independent longitudinal relationship between resting-state EEG alpha asymmetry and behavioral false belief understanding from infancy to early childhood. Findings from Study 1 provide evidence of brain-behavior connections related to false belief understanding in children younger than 3 years of age. Studies 2 and 3 examined task-independent neural correlates of false belief understanding, specifically exploring whether resting-state EEG alpha asymmetry could serve as an early neural marker to predict subsequent behavioral manifestations of false belief understanding at age 4. These three studies reveal potential neural correlates of false belief understanding in toddlerhood and suggest that early-appearing task-independent brain activities may support later false belief understanding ability.

This final chapter synthesizes the dissertation's findings and aims to answer the research questions posited in Section 1.6.1. In Section 4.1, the primary results of the studies are concisely summarized. Section 4.2 discusses these findings in relation to the neural correlates of false belief understanding from infancy to early childhood. Section 4.3 discusses the theoretical implications. Section 4.4 outlines the research limitations and proposes future research directions. The discussion concludes with a summary in Section 4.5.

4.1. Summary of the three studies

In the traditional developmental view, a milestone of Theory of Mind is achieved around age 4, when children start passing the traditional explicit false belief tasks. However, recent research involving explicit false belief tasks with reduced cognitive demands has shown that even before their third birthday, toddlers display correct behavioral prediction of an agent who acts based on a false belief (Grosso et al., 2019; Setoh et al., 2016; Sodian et al., 2024). Furthermore, studies using spontaneous-response paradigms suggested that toddlers under 2 years old and infants as young as 1 year of age can make action predictions based on others' beliefs (e.g., Kaltefleiter et al., 2022; Onishi & Baillargeon, 2005; Southgate et al., 2007). If toddlers and even infants are already sensitive to others' false beliefs in low-demands explicit false belief tasks or novel implicit false belief tasks, it is plausible that cortical regions supporting Theory of Mind are functioning as early as in toddlerhood or even infancy. Although the existing literature suggests the presence of a specialized neural system for processing false belief reasoning, further evidence is needed to clarify the existence of such a system, particularly in young children. Key questions remain: What are the neural correlates of false belief understanding in children younger than 3 years? What task-independent neural activities emerge early on and pave the way for later success in behavioral false belief tasks? What are the characteristics of potential early-emerging precursors that correlate with implicit and explicit false belief understanding? Exploring these questions could enhance our understanding of the development of social cognition and the neural underpinnings of Theory of Mind in early childhood.

In Study 1, we tested children cross-sectionally at approximately 34 months of age using a low-demands false belief task (Grosso et al., 2019; Setoh et al., 2016) on the behavioral level, and a standard explicit false belief task on the neural level (Wellman & Bartsch, 1988; Wellman & Liu, 2004), to investigate brain-behavior connections related to false belief understanding in toddlers under 3 years old. In Study 2, we acquired resting-state EEG alpha asymmetry at 34 months, assessed explicit false belief tasks at 4 years, and measured implicit false belief tasks at both time points to study their interrelations longitudinally. To further support the preliminary evidence from Study 2 concerning explicit false belief understanding in a limited cohort, Study 3 analyzed data from an independent longitudinal dataset to test the generality of the relationship between resting-state EEG alpha asymmetry assessed at 14 months and behavioral explicit false belief understanding at age 4. Studies 2 and 3 aim to explore the development of task-independent brain activity during early childhood, particularly from infancy to toddlerhood, and to investigate its potential correlation with behavioral implicit and explicit false belief competence.

In Study 1, a total of 128 toddlers participated in an ERP belief task (Leo's belief task), adapted from the behavioral "explicit false belief task" (Wellman & Bartsch, 1988; Wellman & Liu, 2004), which explicitly informed children about both reality and the protagonist's belief and asked them where the protagonist would look for the target object. Data obtained from 70 toddlers met the criteria for artifact-free ERP segment inclusion. Of these, 45 toddlers who met the criteria for ERP artifact rejection and completed the low-demands false belief task concurrently were included in the final sample for data analysis. We found that a late positive waveform over the occipital electrode sites distinguished between false belief and true belief conditions only in toddlers who passed the low-demands behavioral false belief task. In contrast, a late negative waveform over the slightly right frontocentral electrode sites consistently distinguished between false belief and true belief conditions regardless of low-demands behavioral false belief competence. These findings raise the possibility that a sensitive neural system supporting false belief understanding may emerge early in brain development. Specifically, the late positive waveforms observed over the occipital electrode sites appear to be a potential neural marker for false belief understanding in toddlers.

In Study 2, we employed a longitudinal design to examine the potential relationship between early resting-state EEG alpha asymmetry and behavioral false belief understanding during the critical developmental period traditionally associated with the explicit comprehension of false beliefs, specifically between the ages of 3 and 4 years. Study 2 assessed resting-state EEG alpha asymmetry across frontal and parietal electrode sites at around age 34 months. Since 4-year-olds already start succeeding in explicit false belief tasks, explicit false belief understanding was assessed at age 4. Considering that implicit false belief understanding was presented at a younger age than 3 years, we measured implicit false belief understanding at both 34 months and 4 years old. Additionally, we assessed a battery of standard language skill assessments and executive functions. We found a significant longitudinal correlation between greater right than left frontal activity at 34 months old and better performance in explicit false belief tasks at 4 years of age, which suggests the resting-state right frontal activities may serve as an early-appearing neural marker of children's later explicit false belief understanding. However, greater right than left parietal activity correlated with superior performance on an implicit false belief task only at 34 months old but not at 4 years of age, showing significant cross-sectional but no longitudinal correlations. This implies that resting-state EEG alpha asymmetry in parietal regions may reflect implicit false belief competence in younger children, yet its reliability as a consistent neural marker does not extend throughout early childhood.

Furthermore, Study 3 analyzed data with a more extensive age range from another independent longitudinal dataset to test the generality of the relationship between resting-state EEG alpha asymmetry assessed at 14 months and explicit false belief understanding at age 4. We found that superior explicit false belief understanding at age 4 was predicted by greater right than left frontal activity even at 14 months, a pattern that aligns with the results observed in Study 2. All these significant correlation results were independent of age and other confounding cognitive abilities (i.e., language, executive function). To further substantiate these findings regarding the relationship between right frontal activity and better explicit false belief understanding, we combined the samples from Study 2 and Study 3, yielding a total of *N* = 80 children who completed the resting-state EEG recordings and 54 children who participated in the explicit false belief tasks. Intriguingly, this analysis revealed a pattern consistent with the results observed in Studies 2 and 3, indicating that greater right than left frontal activity at an earlier age predicted superior explicit false belief understanding at age 4.

4.2. Discussion of the findings on neural correlates of false belief understanding from infancy to early childhood

4.2.1. Neural correlates of explicit false belief understanding may emerge earlier than assumed

This dissertation aims to contribute to our understanding of the developmental origins of Theory of Mind in the brain, with a particular focus on the neural correlates of false belief understanding from infancy to early childhood. Explicit elicited-response reasoning about false beliefs develops between ages 3 and 4 (Gonzales et al., 2018, p. 208; Gopnik & Graf, 1988; Wellman et al., 2001; Yu & Wellman, 2024). However, a new low-demands false belief task produced evidence of explicit false belief reasoning already in 2.5-year-olds (Grosso et al., 2019; Kaltefleiter et al., 2021; Setoh et al., 2016) and performance on this task at 33 months was significantly correlated with standard content false belief task performance at 52 months, independent of co-developing cognitive abilities (Sodian et al., 2024). These findings suggest that explicit false belief understanding may emerge earlier than previously assumed and support the view of conceptual continuity in the development of early Theory of Mind.

Conceptual continuity views propose that false belief reasoning emerges early in life and gradually becomes more efficient and more nuanced with age and experience (Baillargeon et al., 2016; Scott & Baillargeon, 2017). They assume that children understand false beliefs already before age 4, but traditional elicited-prediction false belief tasks are subject to greater processing difficulties (Scott & Baillargeon, 2017). In Study 1, we assessed explicit false belief understanding abilities with a reduced processing demands task and the participants' neural response to false beliefs and true beliefs in a novel ERP explicit belief paradigm. Behaviorally, we confirmed that toddlers younger than 3 years old succeed at a traditional false belief task when overall processing demands are reduced, supporting the substantial continuity in false belief understanding from infancy to childhood. Early difficulties with traditional false belief tasks are primarily due to these tasks' heavy processing demands (Bloom & German, 2000; Carruthers, 2013; Roth & Leslie, 1998; Scott & Baillargeon, 2009). In addition to above-chance performance in the explicit low-demands false belief task, we found corresponding brain-behavior connections in children younger than 3. Our findings revealed that occipital late waveforms differentiated between false belief and true belief conditions only in toddlers who demonstrated behavioral false belief understanding competence. These findings add to the literature (Liu et al., 2009b) suggesting that, even at around 33 to 36 months, toddlers possess a sensitive neural system associated with false belief processing, and the corresponding neural response had a significant connection to their behavioral false belief competence. The connection between neural response to false beliefs and behavioral competence in the traditional false belief tasks speaks to substantial continuity in explicit false belief understanding, both on the behavioral and neural levels.

Our reference to "conceptual continuity" pertains to the developmental trajectory of false belief understanding abilities from an earlier age (around 3 years) to the preschool age (4-6 years). Specifically, we propose that if children exhibit false belief understanding abilities on the behavioral level before 36 months, this may suggest an earlier onset of the cognitive mechanisms underlying false belief understanding, which could be reflected in their neural responses. Toddlers aged 33 to 36 months who pass the low-demands behavioral false belief task exhibit distinct late waveform neural patterns over occipital sites when comparing false belief to true belief conditions. This finding suggests that toddlers may have established a neural foundation for false belief understanding comparable to that observed in older children aged 4–6 years. The presence of such neural distinctions in toddlers aligns with their behavioral competence in false belief understanding, highlighting the early development of cognitive processes associated with this ability. These results provide evidence supporting the notion of developmental continuity in conceptual understanding.

The common neural responses across passers and failers over slightly right frontal areas appear to indicate that behavioral failers do not differ significantly from passers on the neural level. The theoretical account assumes that both groups distinguish between the false and true belief conditions on the level of representing an agent as mentally connected to a state of the world. What distinguishes passers and failers on the neural level seems to be a metarepresentational understanding of false belief (i.e., understanding that an agent genuinely believes that X is true while I know that X is false). False belief understanding requires a representational view of the mind because false beliefs are mental misrepresentations of reality. This understanding necessitates a clear distinction between *sense* and *reference*, since they guide actions in the real situation (referent) as if it were a different situation (sense) (see Section 1.1.2 “Meta-representation view”). In this context, the neural correlates of false belief understanding in 33- to 36-month-old toddlers may be viewed as part of the same conceptual system of Theory of Mind, rather than representing a radical conceptual change.

Furthermore, several researchers noted that successful performance on false belief tasks requires not only reasoning about mental states but also the engagement of domain-general cognitive resources, such as reorienting attention and inhibitory control (Bloom & German, 2000; Devaney, 2018; Leslie et al., 2004; Perner & Lang, 1999; Sabbagh et al., 2006). The degree of cortical overlap between reorienting attention and Theory of Mind has been widely debated, with empirical studies producing contradictory findings. Some studies indicate no significant overlap (Scholz et al., 2009), while others suggest substantial overlap (J. P. Mitchell, 2008). In Study 1, we observed distinct patterns of late waveforms over occipital sites across groups and conditions. This led us to preliminarily speculate about the allocation of attentional and cognitive resources during explicit false belief understanding. In the current ERP paradigm, children participated in a specialized location false belief task that explicitly described the protagonist’s true or false beliefs, eliminating the need for participants to infer these beliefs. In both false belief and true belief conditions, participants initially focused on the actual location of the target object (i.e., reality). In the false belief condition, they subsequently reoriented their attention to address the inconsistency between the initial reality and the protagonist’s subsequent belief about the object’s location, integrating this inconsistency over time. In contrast, the true belief condition did not require reorienting attention to any inconsistencies, as the protagonist’s belief was consistent with the initial reality. Our findings indicate that participants who successfully passed the false belief task exhibited higher amplitude positive waveforms over occipital sites in the false belief condition compared to the true belief condition, suggesting greater allocation of cognitive resources—such as attention and working memory—to the false belief condition. This likely reflects enhanced engagement in distinguishing between true and false scenarios. Conversely, participants who failed to pass the task showed no significant differences in waveform amplitude between the false and true belief conditions, as evidenced by a convergence of subtle late waveform amplitudes across both scenarios. The amplitude of these late waveforms among those who failed, when compared to the distinct amplitudes observed among passers, suggests that failers allocate a similar but limited level of cognitive resources across both conditions. This pattern may reflect a reduced ability to effectively allocate cognitive resources and differentiate between true and false beliefs. As ERP amplitude changes are indicative of differences in the magnitude of neural or cognitive responses (Dinteren et al., 2014; Riggins & Scott, 2020), these findings imply that understanding false beliefs is associated with a more flexible allocation of cognitive resources, particularly with respect to reorienting attention to discrepancies between false beliefs and reality. This observation aligns with theories positing that successful belief reasoning involves the ability to shift attention and working memory in order to reconcile conflicting perspectives, underscoring the crucial role of attentional mechanisms in the development of Theory of Mind (Bartholow & Dickter, 2007; Hopfinger & Mangun, 2001). Existing meta-analytic and review studies revealed that co-activation within the TPJ during both attention and mentalizing tasks may indicate a shared cognitive component (Schuwerk et al., 2017), such as contextual updating (Geng & Vossel, 2013) or bottom-up attentional processes (Cabeza et al., 2012). Future research should incorporate additional cognitive assessments, such as measurements of attention and working memory, and explore whether interventions aimed at enhancing attentional control might facilitate earlier or more robust development of false belief understanding in children.

Executive function, particularly inhibitory control, is crucial for children's Theory of Mind abilities (Sabbagh et al., 2006; Sai et al., 2021). In our study, we used the low-demands false belief task by Setoh et al. (2016), which is a location false belief task following the "Max and the chocolate" format. This task provides information on belief formation while decreasing the processing demands of inhibitory control and response generation. In contrast, the ERP false belief task known as "Leo's belief task" specifically examines the causal influence of beliefs on actions, akin to the “Paul and the backpack” paradigm. Understanding the isolated causal relationship between beliefs and actions remains challenging for young children aged 4. This challenge may largely stem from the executive demands imposed by the task, which interact with its conceptual content (see Sodian et al., 2024). Because executive function is closely associated with frontal cortex development, differences in inhibitory control demands between the two false belief tasks may partially explain the lack of a significant correlation between neural responses to false belief in the frontal region and behavioral competence in the low-demands false belief task. Future research should incorporate high-demand false belief tasks that match the inhibitory control demands of the ERP belief task in neural studies with young children, to further clarify the role of inhibitory control in the brain-behavior relationship underlying false belief understanding. Another promising direction is to explore the potential neural correlation between false belief understanding and inhibitory control during toddlerhood. In addition to assessing false belief understanding on the neural level in the current Study 1, future research could also assess inhibitory control using paradigms such as the go/no-go task. For instance, in an fMRI study, Rothmayr et al. (2011) identified significant overlapping activation in regions associated with both belief reasoning and inhibitory control, including the right superior dorsal MPFC, the right TPJ, the dorsal part of the left TPJ, and lateral prefrontal areas. As previously noted in Section 1.4, ERP techniques offer a significant advantage in precisely characterizing the timing of specific neurocognitive processes with high temporal resolution. Therefore, investigating the potential relationship between false belief understanding and inhibitory control at the electrophysiological level remains an intriguing direction for advancing our understanding of Theory of Mind development in children.

Additionally, Study 1 used independent tasks to assess false belief understanding on both the neural and behavioral levels. The behavioral task, designed with low response-generation and inhibitory-control demands, enabled the categorization of toddlers into false belief passers and failers. Concurrently, the ERP belief task provided explicit information about reality and the protagonist's beliefs, eliciting toddlers' neural responses. The deliberate differentiation between the two tasks ensured that the identified brain-behavior connections reflected substantive cognitive processes in false belief understanding rather than superficial task characteristics. However, it can be argued that tasks measuring the same ability should still correlate even when they have different difficulties, which Setoh et al. achieved by asking the two practice where-questions. Hence, the lack of correlation between false belief competence in the ERP belief task and the low-demands false belief task raises questions about whether the two tasks assess the same ability. Firstly, recent findings from a longitudinal behavioral study on false belief understanding (Sodian et al., 2024) demonstrated a predictive relationship between the low-demands false belief task by Setoh et al. (2016) at 33 months and a standard content false belief task, as well as the sum score of the content and location false belief tasks at 52 months, independent of language and executive function. Further results from the study by Sodian et al. (2024) are consistent with the present findings, indicating that the low-demands false belief task and the explicit location false belief task from the Theory of Mind scale by Wellman & Liu (2004) may be unrelated. The location false belief task (referred to as "Paul and the backpack") is of the same type as Leo's belief task in Study 1. The task does not provide explicit information about the sources of the agent's false belief but rather informs participants of what the agent believes—for example, "Paul thinks his mittens are in the backpack" and "Leo thinks (the target object) is in the bucket/box." Furthermore, participants are informed about the state of reality. The test question, "Where will (the agent) look for (the target object)?" specifically assesses understanding of the causal impact of beliefs on actions when the agent's beliefs are counterfactual.

In summary, false belief understanding is not a discrete, all-or-none skill; rather, different false belief tasks assess distinct dimensions of belief comprehension while placing varied demands on general cognitive abilities. The findings from Study 1, which reveal brain-behavior connections in two different types of false belief tasks, suggest that toddlers' understanding of false beliefs may revolve around a conceptual core involving the interplay between perceptual access, knowledge or belief, and action. Moreover, this ability for false belief understanding appears to manifest earlier in development, both in behavioral and neural terms, than previously assumed.

4.2.2. Task-independent EEG alpha asymmetry may be a stable precursor to the development of explicit, rather than implicit, false belief understanding

Building on the task-dependent neural responses to false and true beliefs observed in Study 1, we extended our research in Studies 2 and 3 by tracking resting-state EEG activity and false belief reasoning longitudinally over early childhood. Our findings revealed early-emerging neural substrates associated with later explicit false belief understanding. Specifically, frontal resting-state EEG alpha asymmetry measured at 34 months—and even as early as 14 months—predicted behavioral explicit false belief understanding at around age 4. These findings align with previous studies on the reliability of frontal resting-state EEG alpha asymmetry. Earlier research has demonstrated stable frontal alpha asymmetry in infants and children across a period of nearly six years (N. A. Fox et al., 1992; Jones et al., 1997; Kim & Bell, 2006; Müller et al., 2015; Vuga et al., 2006). However, these previous studies did not explore the potential relationship between the longitudinal stability of frontal alpha asymmetry and socio-cognitive abilities. The present study is the first to investigate the potential stability of alpha asymmetry in conjunction with behavioral false belief competence during early infancy and toddlerhood. Notably, the significant longitudinal correlation between right frontal alpha asymmetry and explicit false belief understanding persists even after controlling for age, language, and executive function.

The longitudinal relationship between resting-state right frontal activity and 4-year-olds' explicit false belief understanding may arise from its association with other early-emerging socio-cognitive abilities. It is plausible that other aspects of Theory of Mind, which develop alongside frontal activity, serve as precursor abilities to explicit false belief understanding around the age of 4 (Brooks & Meltzoff, 2015; Kloo et al., 2021; Sodian et al., 2016; Yeung et al., 2019). For example, previous studies found that significant aspects of Theory of Mind, such as understanding self versus others (Kaltefleiter et al., 2021; Rochat & Striano, 2002) and grasping others' goals and intentions (Wellman et al., 2004, 2008; Yamaguchi et al., 2009), which manifest by age 2, are underpinned by neural activity in the frontal regions. Some longitudinal studies found that infants' early understanding of intentional action predicts their later explicit false belief competence (Kaltefleiter et al., 2022; Sodian et al., 2016; Wellman et al., 2004, 2008; Yamaguchi et al., 2009; Yott & Poulin-Dubois, 2016), suggesting that early perception and later reasoning about other people may rely on common cognitive mechanisms. Furthermore, previous studies also found that right frontal activity is linked to enhanced perspective-taking (Tullett et al., 2012) and cognitive empathy (Rueckert & Naybar, 2008), both of which are closely related to the representational Theory of Mind.

Additionally, we found a stable longitudinal relationship between resting-state EEG alpha asymmetry and explicit false belief understanding only for frontal alpha asymmetry but not for parietal alpha asymmetry. In other task-dependent neural studies, researchers usually found that right TPJ activity is associated with explicit false belief understanding. The absence of a significant correlation between resting-state parietal asymmetry and explicit false belief understanding in our research may be due to the limited sample size or the potential instability of parietal asymmetry (Müller et al., 2015). Given the small sample size, we were unable to draw a definite conclusion. One possibility is that the results suggest the unreliability of parietal resting-state EEG alpha asymmetry as a neural marker of explicit false belief competence. Supporting this hypothesis, Xiao et al. (2016) found significant developmental changes specifically in the posterior subsystem but not in the frontal subsystem, suggesting distinct developmental trajectories for resting-state neural systems. Further supporting evidence comes from G. Li et al. (2014), who observed developmental changes in parieto-occipital asymmetry. At birth, a small cluster of rightward asymmetry was identified in the parieto-occipital sulcus, which expanded to encompass a larger portion of the parieto-occipital sulcus and cuneus cortex by ages 1 and 2. Another possibility is that task-independent brain activity and task-dependent brain activity should not be considered equivalent. Resting-state EEG typically examines spontaneous neural activity without specific tasks or external stimuli. It reflects the intrinsic functional organization of the brain, including power spectra, and is often used to assess brain states like alertness or pathology. On the other hand, ERPs are time-locked responses to specific sensory, cognitive, or motor events, capturing how the brain processes these stimuli. In this case, it is reasonable to find differences between resting-state EEG right parietal activity and task-dependent ERP or fMRI right parietal activity associated with explicit false belief understanding.

Nevertheless, parietal resting-state EEG alpha asymmetry was found to be cross-sectionally correlated only with implicit false belief understanding at 34 months. However, no longitudinal correlation was found between resting-state EEG right parietal activity at 34 months and implicit false belief understanding at 52 months. On the behavioral level, there is also a cross-sectional correlation between implicit and explicit false belief understanding at 52 months, but no longitudinal correlation between implicit false belief understanding from 34 months to 52 months. Although children’s performance in the implicit false belief task did not improve with age, there may underlie a systematic change in implicit false belief understanding over time rather than an indication of stability (Kaltefleiter et al., 2022). By 52 months, children begin to develop explicit false belief understanding, and it is possible that they use similar strategies to complete implicit tasks at this age, which may be difficult for us to detect accurately. For example, in neurotypical children aged 4 to 8 years, performance in non-verbal anticipatory looking tasks was able to predict that an agent with a false belief about an object’s location will search erroneously for the target object (Meristo et al., 2016).

Furthermore, the observed correlations with linguistic abilities provide additional insights for interpreting our findings. Specifically, we identified a significant association between explicit false belief performance and language, as well as between explicit false belief performance and resting-state EEG alpha asymmetry in frontal regions. In contrast, no significant correlation was found between implicit false belief performance and language. This pattern of correlations suggests the involvement of partly distinct neural mechanisms in implicit and explicit false belief understanding, while also supporting a shared conceptual basis between explicit false belief understanding and language development. It could be argued that performance in explicit false belief tasks is influenced primarily by task-related linguistic demands rather than reflecting Theory of Mind. However, it is crucial to highlight that enhanced performance on explicit false belief tasks, as predicted by greater right frontal activity compared to the left, was independent of age, linguistic abilities, and executive function (Studies 2 and 3). This suggests that while co-developing factors contribute to the development of explicit false belief understanding, they are not solely responsible for advancements in explicit false belief tasks. Moreover, the implicit false belief task showed no correlation with any language assessments or measures of executive function, indicating that early anticipation of an agent’s actions based on a false belief does not rely on these general cognitive abilities. These findings imply that implicit and explicit false belief understanding may be underpinned by distinct neural mechanisms during early development.

The present findings support the hypothesis that frontal alpha asymmetry may serve as a stable trait marker in childhood, whereas parietal alpha asymmetry does not demonstrate similar stability (Harmon-Jones et al., 2010; G. Li et al., 2014; Müller et al., 2015; Xiao et al., 2016). For instance, Müller et al. (2015) reported positive correlations between frontal asymmetry scores measured at 14 and 83 months of age, while Xiao et al. (2016) found evidence of stronger right hemispheric lateralization at age 3, with analyses of intrinsic activity revealing significant developmental reliability in the MPFC subsystem of the default mode network. These findings extend these previous neuroimaging studies that focused on isolated developmental characteristics of the resting-state neural system to their implications for socio-cognitive abilities, particularly false belief understanding. Resting-state EEG alpha asymmetry is influenced by multiple mechanisms along the rostral-caudal plane, and these asymmetries predict task performance in a manner consistent with lesion and neuroimaging studies (Hoptman & Davidson, 1998). Such methods may serve as a valuable tool for assessing neural correlates of cognition in populations for whom longer, more involved methods (e.g., fMRI) are less feasible, such as infants and toddlers. Our results indicate that individual differences in false belief competence during early childhood (approximately 4 years of age) can be predicted by resting-state EEG frontal alpha asymmetry as early as 14 and 34 months of age. Therefore, employing frontal alpha asymmetry measures to explore the processes underlying false belief understanding in childhood may further enhance our understanding of resting-state EEG as a potential trait marker in the development of Theory of Mind.

In summary, in addition to the task-dependent ERP technique discussed above, the task-independent resting-state EEG technique can also enhance our understanding of the development of children’s false belief understanding.

Although we identified a potential neural marker—frontal alpha asymmetry in resting-state EEG—that predicts explicit false belief understanding over a span of more than three years in two independent datasets, the influence of age, language, and executive function was also considered. Nevertheless, due to the limited sample size, these findings should be regarded as a preliminary exploration in this field.

4.3. Theoretical implications

Integrating behavioral false belief competence with electrophysiological neural responses during early childhood informs theoretical perspectives on early Theory of Mind development. Proponents of conceptual continuity theories propose that false belief reasoning develops from infancy through a single underlying system, and that infants' reasoning is qualitatively similar to that of older children and adults (Baillargeon et al., 2016; Scott & Baillargeon, 2017). They posit that young children struggle with traditional false belief tasks primarily due to increased processing demands, which impede their performance (Scott et al., 2022; Scott & Baillargeon, 2017). The findings from Study 1, demonstrating brain-behavior connections in low-demands false belief competence in 33- to 36-month-old toddlers, support this account. Specifically, the significant correlation between toddlers' behavioral false belief competence in a low-demands false belief task and their neural responses to belief information in an independent ERP task bolsters conceptual continuity claims. This indicates that behavioral explicit false belief competence might be present from an early age and has corresponding brain-behavior connections associated with false belief processing.

The longitudinal relationships found in Studies 2 and 3 between greater right frontal activity in infants and toddlers during resting-state EEG recordings and behavioral explicit false belief competence at age 4 align with the conceptual continuity view but may also support a developmental enrichment view. Regarding the longitudinal relationships between resting-state EEG alpha asymmetry and explicit false belief understanding, future studies should investigate whether Theory of Mind precursor abilities—such as perspective-taking and metacognition (Kaltefleiter et al., 2021)—that develop during infancy support later explicit false belief understanding. Examining these precursor abilities may help future research elucidate the development of false belief reasoning during early childhood.

Moreover, the observed cross-sectional relationships between greater parietal activity and implicit false belief understanding at age 3, along with the associations between implicit and explicit false belief reasoning at age 4, may indicate developmental changes in children's response to the implicit false belief stimuli between ages 3 and 4. By age 4, children begin to develop explicit false belief understanding, and it is possible that they use strategies similar to explicit reasoning to complete implicit tasks at this age, which may not be accurately detected. In this context, implicit false belief understanding may undergo a functional shift in specialization between ages 3 and 4. However, we did not find overlapping characteristics of resting-state EEG alpha asymmetry between implicit and explicit false belief understanding. This suggests that, at early age ranges, implicit and explicit false belief understanding may develop based on partly distinct neural mechanisms, supporting the two-system accounts. For subsequent developmental stages after children develop explicit false belief understanding, whether implicit and explicit false belief understanding share overlapping neural underpinnings or develop distinctly remains to be investigated in future studies.

4.4. Limitations and suggestions for future research

A constraint impacting the research discussed in this dissertation was the relatively small sample size. In Study 1, although the original sample size for the ERP task was 128 toddlers who completed Leo's belief task, and the original sample size for the behavioral low-demands false belief task was 75 toddlers who completed Lily's belief task, we ultimately obtained data from only

45 toddlers who completed both the ERP belief task and met the artifact-free criteria for pre-processing, while also completing the low-demands false belief task. In Studies 2 and 3, while the original sample size for behavioral false belief tasks in the Crossing project and TOMEC/TOMII project was over 80 participants, fewer than 30 participants completed both the resting-state EEG recording and the behavioral false belief tasks. Although the exclusion rate in Study 1 was 41.18%, which is similar to other EEG studies with young children (Hoehl & Wahl, 2012; Stets et al., 2012), and we combined data from Studies 2 and 3 to form a larger sample to mitigate concerns about small sample sizes, we must acknowledge the significant data loss due to high artifact rates in the electrophysiological data and low participation in the behavioral data. Despite our efforts, these factors represented a limitation that should be considered when interpreting the findings. Future research could explore the integration of offline laboratory methods with remote web-based tools for testing toddlers and infants, thereby enhancing the recruitment of larger and more diverse samples (Kaletsch & Liszkowski, 2024; Steffan et al., 2024).

In Study 1, we recruited only 33- to 36-month-old toddlers to complete false belief tasks at both the behavioral and neural levels, allowing us to examine the neural correlates of false belief understanding in toddlers younger than 3 years old. However, to gain a more comprehensive developmental picture of infants' and toddlers' neural correlates of false belief competence, we should systematically explore the neural correlates of false belief understanding across a wider age range, including at age 4, using both behavioral and neural measures. The current finding that behavioral passers below 36 months exhibited neural distinctions over occipital sites, whereas 4- to 6-year-olds in Liu et al.'s (2009b) study showed corresponding distinctions over frontal sites, warrants further investigation. It is important to note that the behavioral false belief task used by Liu et al. (2009b) was not a low-demands task like that of Setoh et al. (2016). Therefore, the correlation between behavioral competence and frontal neural responses in 4- to 6-year-olds may reflect an interaction between inhibitory control (and other cognitive demands) and belief understanding rather than belief understanding alone. Future research should explore systematic relationships between neural responses to belief information and the varying demands of false belief tasks (i.e., low-demands versus high-demand) around age 4. This would help clarify developmental differences in the performance of 3- and 4-year-olds in low-demands and high-demand false belief tasks, providing valuable insights into whether explicit false belief reasoning undergoes substantial developmental changes and how processing demands influence the neural correlates of false belief competence.

Further research should also extend beyond the neural correlates of false belief understanding to examine Theory of Mind precursor abilities. This dissertation explored various types of false belief tasks, including explicit (i.e., low-demands false belief, location false belief, content false belief) and implicit false belief tasks. However, a more comprehensive approach to understanding how children develop mental state reasoning should also investigate precursor abilities, such as goal-directed attention, visual perception, and desire reasoning on the neural level. Previous research has demonstrated a link between early goal-directed attention in infancy and later Theory of Mind abilities (Aschersleben et al., 2008; Peterson et al., 2005; Wellman, 2002; Wellman et al., 2004, 2008; Wellman & Liu, 2004; Yamaguchi et al., 2009). For example, a significant correlation was found between 6-month-olds' attention to goal-directed action and their ability to solve false belief tasks at age four, independent of language abilities (Aschersleben et al., 2008). Similarly, 14-month-olds' habituation to human intentional action significantly predicted later mentalistic reasoning in preschool (Wellman et al., 2004), and 11-month-olds' social attention predicted their later false belief understanding, highlighting a continuity in social cognition (Wellman et al., 2008). In this dissertation, we found significant relationships between early resting-state EEG alpha asymmetry and later false belief understanding. However, it remains unclear whether early Theory of Mind precursor abilities contribute to the observed EEG asymmetry and, if so, how these abilities might shape the relationship between early EEG markers and later false belief competence. Understanding how neural and behavioral correlates of beliefs and other Theory of Mind precursor abilities interact will provide valuable insights into the development of a fully formed Theory of Mind.

Moreover, when comparing the findings from Study 1 with those from subsequent Studies 2 and 3, we observe that all studies identified right frontal neural activity associated with false belief understanding—either elicited by false belief stimuli (Study 1) or observed in the absence of external stimuli (Studies 2 and 3). Specifically, Study 1 found a late negative waveform over slightly right-lateralized frontocentral electrode sites (e.g., FC2, FC6, C4) that consistently differentiated between false belief and true belief conditions, independent of performance on low-demands behavioral false belief tasks. However, it is not possible to determine whether this neural activity reflects a non-mentalistic match–mismatch detection process or a non-representational mental state attribution process occurring over the right frontocentral electrode sites. Combined with the finding that early-appearing resting-state EEG right frontal activity predicts later explicit false belief understanding at age 4, the potential relationship between task-dependent frontal activity and task-independent frontal activity remains unclear. How these neural processes synergize to facilitate the development of false belief understanding is an intriguing question for future research.

4.5. Conclusion

In summary, our results demonstrate neural correlates of false belief understanding from infancy to early childhood, utilizing both task-dependent and task-independent electrophysiological methods. Together with previous behavioral findings (e.g., Grosso et al., 2019; Setoh et al., 2016; Sodian et al.,

2024) and neural findings (e.g., Grosse Wiesmann et al., 2020; Liu et al., 2009b; Richardson et al., 2018) associated with false belief understanding in early childhood, our findings reveal:

1) The existence of a sensitive neural system over posterior regions for explicit false belief understanding, even in toddlers under 3 years old (Study 1).

2) Resting-state EEG alpha asymmetry over frontal regions shows reliable individual neural differences over time that relate to behavioral explicit false belief competence, and early measurements of this marker predict subsequent explicit false belief understanding ability at 4 years of age (Studies 2 and 3).

Given that this research is a preliminary exploration of the developmental origins of Theory of Mind in the brain during early childhood using electrophysiological techniques, future studies should recruit larger sample sizes to replicate these findings. Further investigations should explore the development of false belief understanding—including both explicit and implicit forms—in infants and toddlers across multiple time points. Moreover, future studies should examine the underlying psychological mechanisms and related cognitive processes to improve our understanding of the early development of Theory of Mind abilities.

References

Aichhorn, M., Perner, J., Kronbichler, M., Staffen, W., & Ladurner, G. (2006). Do visual perspective tasks need theory of mind? *NeuroImage*, *30*(3), 1059–1068. https://doi.org/10.1016/j.neuroimage.2005.10.026

Aichhorn, M., Perner, J., Weiss, B., Kronbichler, M., Staffen, W., & Ladurner, G. (2009). Temporo-parietal junction activity in theory-of-mind tasks: Falseness, beliefs, or attention. *Journal of Cognitive Neuroscience*, *21*(6), 1179–1192. https://doi.org/10.1162/jocn.2009.21082

Allen, J. J. B., Urry, H. L., Hitt, S. K., & Coan, J. A. (2004). The stability of resting frontal electroencephalographic asymmetry in depression. *Psychophysiology*, *41*(2), 269–280. https://doi.org/10.1111/j.1469-8986.2003.00149.x

Amft, M., Bzdok, D., Laird, A. R., Fox, P. T., Schilbach, L., & Eickhoff, S. B. (2015). Definition and characterization of an extended social-affective default network. *Brain Structure and Function*, *220*(2), 1031–1049. https://doi.org/10.1007/s00429-013-0698-0

Amodio, D. M., & Frith, C. D. (2006). Meeting of minds: The medial frontal cortex and social cognition. *Nature Reviews Neuroscience*, *7*(4), 268–277. https://doi.org/10.1038/nrn1884

Andrews-Hanna, J. R., Saxe, R., & Yarkoni, T. (2014). Contributions of episodic retrieval and mentalizing to autobiographical thought: Evidence from functional neuroimaging, resting-state connectivity, and fMRI meta-analyses. *NeuroImage*, *91*, 324–335. https://doi.org/10.1016/j.neuroimage.2014.01.032

Apperly, I. A., & Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review*, *116*(4), 953–970. https://doi.org/10.1037/a0016923

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. https://doi.org/10.1162/0898929042947928

Apperly, I. A., Warren, F., Andrews, B. J., Grant, J., & Todd, S. (2011). Developmental continuity in theory of mind: Speed and accuracy of belief–desire reasoning in children and adults. *Child Development*, *82*(5), 1691–1703. https://doi.org/10.1111/j.1467-8624.2011.01635.x

Aschersleben, G., Hofer, T., & Jovanovic, B. (2008). The link between infant attention to goal-directed action and later theory of mind abilities. *Developmental Science*, *11*(6), 862–868. https://doi.org/10.1111/j.1467-7687.2008.00736.x

Astington, J. W., & Baird, J. A. (Eds.). (2005). *Why Language Matters for Theory of Mind*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195159912.001.0001

Astington, J. W., Harris, P. L., & Olson, D. R. (1988). *Developing theories of mind*. Cambridge University Press.

Aykan, S., Kalaycioğlu, C., & Nalcaci, E. (2021). Alpha coherence decreases in the right posterior cortex reflects better theory of mind ability. *Psychiatry and Behavioral Sciences*, *11*, 63. https://doi.org/10.5455/PBS.20201125074821

Bailey, K., Mlynarczyk, G., & West, R. (2016). Slow wave activity related to working memory maintenance in the N-back task. *Journal of Psychophysiology*, *30*(4), 141–154. https://doi.org/10.1027/0269-8803/a000164

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. https://doi.org/10.1016/j.cogdev.2018.06.001

Baillargeon, R., Scott, R. M., & Bian, L. (2016). Psychological reasoning in infancy. *Annual Review of Psychology*, *67*(1), 159–186. https://doi.org/10.1146/annurev-psych-010213-115033

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

Baillargeon, R., Scott, R. M., He, Z., Sloane, S., Setoh, P., Jin, K., Wu, D., & Bian, L. (2015). Psychological and sociomoral reasoning in infancy. In M. Mikulincer, P. R. Shaver, E. Borgida, & J. A. Bargh (Eds.), *APA handbook of personality and social psychology, Volume 1: Attitudes* *and social cognition.* (pp. 79–150). American Psychological Association. https://doi.org/10.1037/14341-003

Baker, S. T., Leslie, A. M., Gallistel, C. R., & Hood, B. M. (2016). Bayesian change-point analysis reveals developmental change in a classic theory of mind task. *Cognitive Psychology*, *91*, 124–149. https://doi.org/10.1016/j.cogpsych.2016.08.001

Balaban, N., Friedmann, N., & Ziv, M. (2016). Theory of mind impairment after right-hemisphere damage. *Aphasiology*, *30*(12), 1399–1423. https://doi.org/10.1080/02687038.2015.1137275

Bardi, L., Desmet, C., Nijhof, A., Wiersema, J. R., & Brass, M. (2016). Brain activation for spontaneous and explicit false belief tasks overlaps: New fMRI evidence on belief processing and violation of expectation. *Social Cognitive and Affective Neuroscience*, nsw143. https://doi.org/10.1093/scan/nsw143

Baron-Cohen, S. (2000). Theory of mind and autism: A review. In *International review of research in mental retardation* (Vol. 23, pp. 169–184). Academic Press. https://doi.org/10.1016/S0074-7750(00)80010-5

Baron-Cohen, S. (2001). Theory of mind in normal development and autism. *Prisme*, *34*.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a “theory of mind”? *Cognition*, *21*(1), 37–46. https://doi.org/10.1016/0010-0277(85)90022-8

Baron-Cohen, S., Ring, H., Moriarty, J., Schmitz, B., Costa, D., & Ell, P. (1994). Recognition of mental state terms: Clinical findings in children with autism and a functional neuroimaging study of normal adults. *The British Journal of Psychiatry : The Journal of Mental Science*, *165*, 640–649. https://doi.org/10.1192/bjp.165.5.640

Barriga-Paulino, C. I., Rodríguez-Martínez, E. I., Rojas-Benjumea, M. Á., & Gómez, C. M. (2014). Slow wave maturation on a visual working memory task. *Brain and Cognition*, *88*, 43–54. https://doi.org/10.1016/j.bandc.2014.04.003

Bartholow, B. D., & Dickter, C. L. (2007). Social cognitive neuroscience of person perception: A selective review focused on the event-related brain potential. In E. Harmon-Jones & P. Winkielman (Eds.), *Social*

Neuroscience: Integrating Biological and Psychological Explanations of Social Behavior (pp. 376–400). Guilford Press.

Bartsch, K. (1996). Between desires and beliefs: Young children's action predictions. *Child Development*, *67*(4), 1671. https://doi.org/10.2307/1131724

Bartsch, K., & Wellman, H. M. (1995). *Children talk about the mind*. Oxford university press.

Beaudoin, C., Leblanc, É., & Beauchamp, M. H. (2020). Systematic review and inventory of theory of mind measures for young children. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.02905

Bell, M. A. (2001). Brain electrical activity associated with cognitive processing during a looking version of the a-not-b task. *Infancy*, *2*, 311–330.

Bell, M. A. (2002). Power changes in infant EEG frequency bands during a spatial working memory task. *Psychophysiology*, *39*, 450–458.

Bell, M. A., & Fox, N. A. (1992). The relations between frontal brain electrical activity and cognitive development during infancy. *Child Development*, *63*, 1142–1163.

Bell, M. A., & Fox, N. A. (1997). Individual differences in object permanence performance at 8 months: Locomotor experience and brain electrical activity. *Developmental Psychobiology*, *31*, 287–297.

Bennett, J. (1978). Some remarks about concepts. *Behavioral and Brain Sciences*, *1*(4), 557–560. Cambridge Core. https://doi.org/10.1017/S0140525X00076573

Bergman, L. R., Edklund, R., Magnusson, D., Magnusson, D., & Bergman, L. R. (1989). Studying individual development: Problems and methods. In *Problems and Methods in Longitudinal Research: Stability and Change*. Cambridge University Press.

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. https://doi.org/10.1016/S0010-0277(00)00096-2

Boccadoro, S., Cracco, E., Hudson, A. R., Bardi, L., Nijhof, A. D., Wiersema, J. R., Brass, M., & Mueller, S. C. (2019). Defining the neural correlates of spontaneous theory of mind (ToM): An fMRI multi-study investigation. *NeuroImage*, *203*, 116193. https://doi.org/10.1016/j.neuroimage.2019.116193

Boustani, N. (2023). How does the brain support theory of mind in children an EEG/ERP review. *Journal of Cognition, Emotion and Education*, *1*(1). https://doi.org/10.22034/cee.2023.171194

Bowman, L. C., Dodell-Feder, D., Saxe, R., & Sabbagh, M. A. (2019). Continuity in the neural system supporting children’s theory of mind development: Longitudinal links between task-independent EEG and task-dependent fMRI. *Developmental Cognitive Neuroscience*, *40*, 100705. https://doi.org/10.1016/j.dcn.2019.100705

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. https://doi.org/10.1111/j.1467-7687.2012.01158.x

Brooker, R. J., Bates, J. E., Buss, K. A., Canen, M. J., Dennis-Tiwary, T. A., Gatzke-Kopp, L. M., Hoyniak, C., Klein, D. N., Kujawa, A., Lahat, A., Lamm, C., Moser, J. S., Petersen, I. T., Tang, A., Woltering, S., & Schmidt, L. A. (2020). Conducting event-related potential (ERP) research with young children: A review of components, special considerations, and recommendations for research on cognition and emotion. *Journal of Psychophysiology*, *34*(3), 137–158. https://doi.org/10.1027/0269-8803/a000243

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. https://doi.org/10.1016/j.jecp.2014.09.010

Buckner, R. L., Andrews-Hanna, J. R., & Schacter, D. L. (2008). The brain’s default network. *Annals of the New York Academy of Sciences*, *1124*(1), 1–38. https://doi.org/10.1196/annals.1440.011

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. https://doi.org/10.1016/j.cognition.2009.05.006

Buttelmann, D., Over, H., Carpenter, M., & Tomasello, M. (2014). Eighteen-month-olds understand false beliefs in an unexpected-contents task. *Journal of Experimental Child Psychology*, *119*, 120–126.

Butterfill, S. A., & Apperly, I. A. (2013). How to construct a minimal theory of mind. *Mind & Language*, *28*(5), 606–637. https://doi.org/10.1111/mila.12036

Cabeza, R., Ciaramelli, E., & Moscovitch, M. (2012). Cognitive contributions of the ventral parietal cortex: An integrative theoretical account. *Trends in Cognitive Sciences*, *16*(6), 338–352. https://doi.org/10.1016/j.tics.2012.04.008

Cao, B., Li, Y., Li, F., & Li, H. (2012). Electrophysiological difference between mental state decoding and mental state reasoning. *Brain Research*, *1464*, 53–60. https://doi.org/10.1016/j.brainres.2012.05.009

Carey, S. (1985). *Conceptual change in childhood*. MIT Press.

Carlson, S. M., Koenig, M. A., & Harms, M. B. (2013a). Theory of mind. *WIREs Cognitive Science*, *4*(4), 391–402. https://doi.org/10.1002/wcs.1232

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. (2001). Individual Differences in Inhibitory Control and Children’s Theory of Mind. *Child Development*, *72*(4), 1032–1053. https://doi.org/10.1111/1467-8624.00333

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*, *11*(2), 73–92. https://doi.org/10.1002/icd.298

Carlson, S. M., Zelazo, P. D., & Faja, S. (2013b). Executive function. In P. D. Zelazo (Ed.), *The Oxford Handbook of Developmental Psychology, Vol. 1* (pp. 705–743). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199958450.013.0025

Carrington, S. J., & Bailey, A. J. (2008). Are there theory of mind regions in the brain? A review of the neuroimaging literature. *Human Brain Mapping*, *30*(8), 2313–2335. https://doi.org/10.1002/hbm.20671

Carruthers, P. (2013). Mindreading in infancy. *Mind & Language*, *28*(2), 141–172. https://doi.org/10.1111/mila.12014

Carruthers, P. (2016). Two systems for mindreading? *Review of Philosophy and Psychology*, *7*(1), 141–162. https://doi.org/10.1007/s13164-015-0259-y

Chen, L., Cheung, H., Szeto, C.-Y., Zhu, Z., & Wang, S. (2012). Do false belief and verb non-factivity share similar neural circuits? *Neuroscience Letters*, *510*(1), 38–42. https://doi.org/10.1016/j.neulet.2011.12.065

Clements, W. A., & Perner, J. (1994). Implicit understanding of belief. *Cognitive Development*, *9*(4), 377–395. https://doi.org/10.1016/0885-2014(94)90012-4

Contreras, J. M., Schirmer, J., Banaji, M. R., & Mitchell, J. P. (2013). Common brain regions with distinct patterns of neural responses during mentalizing about groups and individuals. *Journal of Cognitive Neuroscience*, *25*(9), 1406–1417. https://doi.org/10.1162/jocn_a_00403

Corradi-Dell'Acqua, C., Hofstetter, C., & Vuilleumier, P. (2014). Cognitive and affective theory of mind share the same local patterns of activity in posterior temporal but not medial prefrontal cortex. *Social Cognitive and Affective Neuroscience*, *9*(8), 1175–1184. https://doi.org/10.1093/scan/nst097

Crivello, C., & Poulin-Dubois, D. (2018). Infants' false belief understanding: A non-replication of the helping task. *Cognitive Development*, *46*, 51–57.

De Bruin, L. C., & Newen, A. (2012). An association account of false belief understanding. *Cognition*, *123*(2), 240–259.

de Villiers, J. G., & de Villiers, P. A. (2003). Language for thought: Coming to understand false beliefs. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and thought* (pp. 335–384). MIT Press.

de Villiers, J. G., & de Villiers, P. A. (2014). The role of language in theory of mind development. *Topics in Language Disorders*, *34*(4), 313–328. https://doi.org/10.1097/TLD.0000000000000037

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. https://doi.org/10.1016/S0885-2014(02)00073-4

DeBoer, T., Scott, L. S., & Nelson, C. A. (2007). Methods for acquiring and analyzing infant event-related potentials. In M. de Haan (Ed.), *Infant EEG and event-related potentials* (pp. 5–37). Psychology Press.

Deco, G., Jirsa, V. K., & McIntosh, A. R. (2011). Emerging concepts for the dynamical organization of resting-state activity in the brain. *Nature Reviews Neuroscience*, *12*(1), 43–56. https://doi.org/10.1038/nrn2961

Dennett, D. C. (1978). Beliefs about beliefs. *Behavioral and Brain Sciences*, *1*(4), 568–570. Cambridge Core. https://doi.org/10.1017/S0140525X00076664

Deoni, S. C. L., Dean, D. C., Remer, J., Dirks, H., & O'Muircheartaigh, J. (2015). Cortical maturation and myelination in healthy toddlers and young children. *NeuroImage*, *115*, 147–161. https://doi.org/10.1016/j.neuroimage.2015.04.058

Devaney, K. (2018). *Functional MRI investigations of human temporoparietal junction: Attention, response inhibition, theory of mind, and long-term meditation effects*. Boston University.

Devine, R. T., & Hughes, C. (2014). Relations between false belief understanding and executive function in early childhood: A meta-analysis. *Child Development*, *85*(5), 1777–1794.

Dien, J. (1998). Issues in the application of the average reference: Review, critiques, and recommendations. *Behavior Research Methods, Instruments, & Computers*, *30*(1), 34–43. https://doi.org/10.3758/BF03209414

Dienes, Z., & Perner, J. (1999). A theory of implicit and explicit knowledge. *Behavioral and Brain Sciences*, *22*(5), 735–808. https://doi.org/10.1017/S0140525X99002186

Dinteren, R. van, Arns, M., Jongsma, M. L. A., & Kessels, R. P. C. (2014). P300 development across the lifespan: A systematic review and meta-analysis. *PLOS ONE*, *9*(2), e87347. https://doi.org/10.1371/journal.pone.0087347

Dodell-Feder, D., Koster-Hale, J., Bedny, M., & Saxe, R. (2011). fMRI item analysis in a theory of mind task. *NeuroImage*, *55*(2), 705–712. https://doi.org/10.1016/j.neuroimage.2010.12.040

Doherty, M. J., & Perner, J. (2020). Mental files: Developmental integration of dual naming and theory of mind. *Developmental Review*, *56*, 100909. https://doi.org/10.1016/j.dr.2020.100909

Döhnel, K., Schuwerk, T., Meinhardt, J., Sodian, B., Hajak, G., & Sommer, M. (2012). Functional activity of the right temporo-parietal junction and of the medial prefrontal cortex associated with true and false belief reasoning. *NeuroImage*, *60*(3), 1652–1661. https://doi.org/10.1016/j.neuroimage.2012.01.073

Donchin, E., & Coles, M. G. H. (1988). Is the P300 component a manifestation of context updating? *Behavioral and Brain Sciences*, *11*(03), 357. https://doi.org/10.1017/S0140525X00058027

Dörrenberg, S., Rakoczy, H., & Liszkowski, U. (2018). How (not) to measure infant theory of mind: Testing the replicability and validity of four non-verbal measures. *Cognitive Development*, *46*, 12–30.

Dugué, L., Merriam, E. P., Heeger, D. J., & Carrasco, M. (2018). Specific visual subregions of TPJ mediate reorienting of spatial attention. *Cerebral Cortex*, *28*(7), 2375–2390. https://doi.org/10.1093/cercor/bhx140

Elbert, T., Heim, S., & Rockstroh, B. (2001). Neural plasticity and development. Ch 14. In C. A. Nelson & M. Luciana (Eds.), *Handbook of developmental cognitive neuroscience*. MIT Press.

Elsner, B., Jeschonek, S., & Pauen, S. (2013). Event-related potentials for 7-month-olds’ processing of animals and furniture items. *Developmental Cognitive Neuroscience*, *3*, 53–60. https://doi.org/10.1016/j.dcn.2012.09.002

Evans, J. St. B. T., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science*, *8*(3), 223–241. https://doi.org/10.1177/1745691612460685

Falck-Ytter, T., Gredebäck, G., & Von Hofsten, C. (2006). Infants predict other people’s action goals. *Nature Neuroscience*, *9*(7), 878–879. https://doi.org/10.1038/nn1729

Federmeier, K. D., & Laszlo, S. (2009). Time for meaning: Electrophysiology provides insights into the dynamics of representation and processing in semantic memory. In D. H. Ross (Ed.), *Psychology of Learning and Motivation—Advances in Research and Theory* (Vol. 51, pp. 1–44). Elsevier. https://doi.org/10.1016/S0079-7421(09)51001-8

Filippi, C., Choi, Y. B., Fox, N. A., & Woodward, A. L. (2020). Neural correlates of infant action processing relate to theory of mind in early childhood. *Developmental Science*, *23*(2). https://doi.org/10.1111/desc.12876

Fizke, E., Butterfill, S., Van De Loo, L., Reindl, E., & Rakoczy, H. (2017). Are there signature limits in early theory of mind? *Journal of Experimental Child Psychology*, *162*, 209–224. https://doi.org/10.1016/j.jecp.2017.05.005

Flanagan, J. R., & Johansson, R. S. (2003). Action plans used in action observation. *Nature*, *424*(6950), 769–771. https://doi.org/10.1038/nature01861

Flavell, J. H. (2000). Development of children's knowledge about the mental world. *International Journal of Behavioral Development*, *24*(1), 15–23. https://doi.org/10.1080/016502500383421

Flavell, J. H., Flavell, E. R., & Green, F. L. (1983). Development of the appearance-reality distinction. *Cognitive Psychology*, *15*(1), 95–120.

Flavell, J. H., Flavell, E. R., Green, F. L., & Moses, L. J. (1990). Young children's understanding of fact beliefs versus value beliefs. *Child Development*, *61*(4), 915–928.

Fletcher, P. C., Happé, F., Frith, U., Baker, S. C., Dolan, R. J., Frackowiak, R. S. J., & Frith, C. D. (1995). Other minds in the brain: A functional imaging study of "theory of mind" in story comprehension. *Cognition*, *57*(2), 109–128. https://doi.org/10.1016/0010-0277(95)00692-R

Fodor, J. A. (1987). *Psychosemantics: The problem of meaning in the philosophy of mind* (Vol. 2). MIT press.

Fogassi, L., Ferrari, P. F., Gesierich, B., Rozzi, S., Chersi, F., & Rizzolatti, G. (2005). Parietal lobe: From action organization to intention understanding. *Science*, *308*(5722), 662–667. https://doi.org/10.1126/science.1106138

Forgács, B., Gervain, J., Parise, E., Csibra, G., Gergely, G., Baross, J., & Király, I. (2020). Electrophysiological investigation of infants' understanding of understanding. *Developmental Cognitive Neuroscience*, *43*, 100783. https://doi.org/10.1016/j.dcn.2020.100783

Forgács, B., Parise, E., Csibra, G., Gergely, G., Jacquey, L., & Gervain, J. (2019). Fourteen-month-old infants track the language comprehension of communicative partners. *Developmental Science*, *22*(2), e12751. https://doi.org/10.1111/desc.12751

Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Van Essen, D. C., & Raichle, M. E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proceedings of the National Academy of Sciences*, *102*(27), 9673–9678. https://doi.org/10.1073/pnas.0504136102

Fox, N. A., Bell, M. A., & Jones, N. A. (1992). Individual differences in response to stress and cerebral asymmetry. *Developmental Neuropsychology*, *8*(2–3), 161–184. https://doi.org/10.1080/87565649209540523

Fox, N. A., Rubin, K. H., Calkins, S. D., Marshall, T. R., Coplan, R. J., Porges, S. W., Long, J. M., & Stewart, S. (1995). Frontal activation asymmetry and social competence at four years of age. *Child Development*, *66*(6), 1770. https://doi.org/10.2307/1131909

Friederici, A. D., Friedrich, M., & Weber, C. (2002). Neural manifestation of cognitive and precognitive mismatch detection in early infancy: *Neuroreport*, *13*(10), 1251–1254. https://doi.org/10.1097/00001756-200207190-00006

Frith, C. D., & Frith, U. (1999). Interacting minds—A biological basis. *Science*, *286*(5445), 1692–1695. https://doi.org/10.1126/science.286.5445.1692

Frith, C. D., & Frith, U. (2006). The neural basis of mentalizing. *Neuron*, *50*(4), 531–534. https://doi.org/10.1016/j.neuron.2006.05.001

Frith, C. D., & Frith, U. (2008). Implicit and explicit processes in social cognition. *Neuron*, *60*(3), 503–510. https://doi.org/10.1016/j.neuron.2008.10.032

Frith, U., Morton, J., & Leslie, A. M. (1991). The cognitive basis of a biological disorder: Autism. *Trends in Neurosciences*, *14*(10), 433–438. https://doi.org/10.1016/0166-2236(91)90041-R

Gallagher, H. L., & Frith, C. D. (2003). Functional imaging of 'theory of mind.' *Trends in Cognitive Sciences*, *7*(2), 77–83. https://doi.org/10.1016/S1364-6613(02)00025-6

Gallagher, H. L., Happé, F., Brunswick, N., Fletcher, P. C., Frith, U., & Frith, C. D. (2000). Reading the mind in cartoons and stories: An fMRI study of 'theory of mind' in verbal and nonverbal tasks. *Neuropsychologia*, *38*(1), 11–21. https://doi.org/10.1016/S0028-3932(99)00053-6

Gao, W., Gilmore, J. H., Shen, D., Smith, J. K., Zhu, H., & Lin, W. (2013). The synchronization within and interaction between the default and dorsal attention networks in early infancy. *Cerebral Cortex*, *23*(3), 594–603. https://doi.org/10.1093/cercor/bhs043

Gao, W., Zhu, H., Giovanello, K. S., Smith, J. K., Shen, D., Gilmore, J. H., & Lin, W. (2009). Evidence on the emergence of the brain's default network from 2-week-old to 2-year-old healthy pediatric subjects. *Proceedings of the National Academy of Sciences*, *106*(16), 6790–6795. https://doi.org/10.1073/pnas.0811221106

Garnham, W. A., & Ruffman, T. (2001). Doesn't see, doesn't know: Is anticipatory looking really related to understanding or belief? *Developmental Science*, *4*(1), 94–100. https://doi.org/10.1111/1467-7687.00153

Gasser, T., Bächer, P., & Möcks, J. (1982). Transformations towards the normal distribution of broad band spectral parameters of the EEG. *Electroencephalography and Clinical Neurophysiology*, *53*(1), 119–124. https://doi.org/10.1016/0013-4694(82)90112-2

Geangu, E., Gibson, A., Kaduk, K., & Reid, V. M. (2013). The neural correlates of passively viewed sequences of true and false beliefs. *Social Cognitive and Affective Neuroscience*, *8*(4), 432–437. https://doi.org/10.1093/scan/nss015

Geng, J. J., & Vossel, S. (2013). Re-evaluating the role of TPJ in attentional control: Contextual updating? *Neuroscience & Biobehavioral Reviews*, *37*(10), 2608–2620. https://doi.org/10.1016/j.neubiorev.2013.08.010

Gerstadt, C. L., Hong, Y. J., & Diamond, A. (1994). The relationship between cognition and action: Performance of children 3–7 years old on a stroop-like day-night test. *Cognition*, *53*(2), 129–153.

Gevins, A. (1998). The future of electroencephalography in assessing neurocognitive functioning. *Electroencephalography and Clinical Neurophysiology*, *106*(2), 165–172. https://doi.org/10.1016/S0013-4694(97)00120-X

Ghrear, S., Baimel, A., Haddock, T., & Birch, S. A. J. (2021). Are the classic false belief tasks cursed? Young children are just as likely as older children to pass a false belief task when they are not required to overcome the curse of knowledge. *PLOS ONE*, *16*(2), e0244141. https://doi.org/10.1371/journal.pone.0244141

Gilbert, S. J., Simons, J. S., Frith, C. D., & Burgess, P. W. (2006). Performance-related activity in medial rostral prefrontal cortex (area 10) during low-demand tasks. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(1), 45–58. https://doi.org/10.1037/0096-1523.32.1.45

Gobbini, M. I., Koralek, A. C., Bryan, R. E., Montgomery, K. J., & Haxby, J. V. (2007). Two takes on the social brain: A comparison of theory of mind tasks. *Journal of Cognitive Neuroscience*, *19*(11), 1803–1814. https://doi.org/10.1162/jocn.2007.19.11.1803

Goel, V., Grafman, J., Sadato, N., & Hallett, M. (1995). Modeling other minds. *NeuroReport-International Journal for Rapid Communications of Research in Neuroscience*, *6*(13), 1741–1746.

Gola, A. A. H. (2012). Mental verb input for promoting children's theory of mind: A training study. *Cognitive Development*, *27*(1), 64–76. https://doi.org/10.1016/j.cogdev.2011.10.003

Gonzales, C. R., Fabricius, W. V., & Kupfer, A. S. (2018). Introspection plays an early role in children's explicit theory of mind development. *Child Development*, *89*(5), 1545–1552. https://doi.org/10.1111/cdev.12876

Goodman, N. D., Baker, C. L., Bonawitz, E. B., Mansinghka, V. K., Gopnik, A., Wellman, H., Schulz, L., & Tenenbaum, J. B. (2006). *Intuitive theories of mind: A rational approach to false belief*. *6*.

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*, *59*(1), 26–37. https://doi.org/10.1111/j.1467-8624.1988.tb03192.x

Gopnik, A., & Graf, P. (1988). Knowing how you know: Young children’s ability to identify and remember the sources of their beliefs. *Child Development*, *59*(5), 1366. https://doi.org/10.2307/1130499

Gopnik, A., & Wellman, H. M. (1992). Why the child’s theory of mind really is a theory. *Mind & Language*, *7*(1–2), 145–171. https://doi.org/10.1111/j.1468-0017.1992.tb00202.x

Gopnik, A., & Wellman, H. M. (1994). The theory theory. In L. A. Hirschfeld & S. A. Gelman (Eds.), *Mapping the Mind: Domain Specificity in Cognition and Culture* (pp. 257–293). Cambridge University Press.

Gopnik, A., & Wellman, H. M. (2012). Reconstructing constructivism: Causal models, bayesian learning mechanisms, and the theory theory. *Psychological Bulletin*, *138*(6), 1085–1108. https://doi.org/10.1037/a0028044

Gratton, G., Coles, M. G. H., & Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology*, *55*(4), 468–484. https://doi.org/10.1016/0013-4694(83)90135-9

Grèzes, J., Frith, C. D., & Passingham, R. E. (2004). Inferring false beliefs from the actions of oneself and others: An fMRI study. *NeuroImage*, *21*(2), 744–750. https://doi.org/10.1016/S1053-8119(03)00665-7

Griffin, R., Friedman, O., Ween, J., Winner, E., Happé, F., & Brownell, H. (2006). Theory of mind and the right cerebral hemisphere: Refining the scope of impairment. *Laterality: Asymmetries of Body, Brain and Cognition*, *11*(3), 195–225. https://doi.org/10.1080/13576500500450552

Grimm, H., Aktaş, M., & Frevert, S. (2015). *Sprachentwicklungstest für drei- bis fünfjährige Kinder*. Hogrefe.

Grosse Wiesmann, C. (2018). *The emergence of theory of mind: Cognitive and neural basis of false belief understanding in preschool age*. Max Planck Institute for Human Cognitive and Brain Sciences.

Grosse Wiesmann, C., Friederici, A. D., Disla, D., Steinbeis, N., & Singer, T. (2018). Longitudinal evidence for 4-year-olds' but not 2- and 3-year-olds' false belief-related action anticipation. *Cognitive Development*, *46*, 58–68. https://doi.org/10.1016/j.cogdev.2017.08.007

Grosse Wiesmann, C., Friederici, A. D., Singer, T., & Steinbeis, N. (2020). Two systems for thinking about others' thoughts in the developing brain. *Proceedings of the National Academy of Sciences*, *117*(12), 6928–6935. https://doi.org/10.1073/pnas.1916725117

Grosse Wiesmann, C., Friederici, A. D., Singer, T., & Steinbeis, N. (2017a). Implicit and explicit false belief development in preschool children. *Developmental Science*, *20*(5), e12445. https://doi.org/10.1111/desc.12445

Grosse Wiesmann, C., Schreiber, J., Singer, T., Steinbeis, N., & Friederici, A. D. (2017b). White matter maturation is associated with the emergence of theory of mind in early childhood. *Nature Communications*, *8*(1), 14692. https://doi.org/10.1038/ncomms14692

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. https://doi.org/10.1016/j.infbeh.2018.09.012

Guan, Y., Keil, A., & Farrar, M. J. (2020). Electrophysiological dynamics of false belief understanding and complementation syntax in school-aged children: Oscillatory brain activity and event-related potentials. *Journal of Experimental Child Psychology*, *198*, 104905. https://doi.org/10.1016/j.jecp.2020.104905

Gweon, H., Dodell-Feder, D., Bedny, M., & Saxe, R. (2012). Theory of mind performance in children correlates with functional specialization of a brain region for thinking about thoughts. *Child Development*, *83*(6), 1853–1868. https://doi.org/10.1111/j.1467-8624.2012.01829.x

Gweon, H., & Saxe, R. (2013). Developmental cognitive neuroscience of theory of mind. In *Neural Circuit Development and Function in the Brain* (pp. 367–377). Elsevier. https://doi.org/10.1016/B978-0-12-397267-5.00057-1

Hagemann, D., Naumann, E., Thayer, J. F., & Bartussek, D. (2002). Does resting electroencephalograph asymmetry reflect a trait? An application of latent state-trait theory. *Journal of Personality and Social Psychology*, *82*(4), 619–641. https://doi.org/10.1037/0022-3514.82.4.619

Happé, F., Briskman, J., & Frith, U. (2001). Exploring the cognitive phenotype of autism: Weak “central coherence” in parents and siblings of children with autism: i. experimental tests. *Journal of Child Psychology and Psychiatry*, *42*(3), 299–307. https://doi.org/10.1111/1469-7610.00723

Happé, F., & Frith, U. (1995). Theory of Mind in Autism. In E. Schopler & G. B. Mesibov (Eds.), *Learning and Cognition in Autism* (pp. 177–197). Springer US. https://doi.org/10.1007/978-1-4899-1286-2_10

Harman, G. (1978). Studying the chimpanzee’s theory of mind. *Behavioral and Brain Sciences*, *1*(4), 576–577. Cambridge Core. https://doi.org/10.1017/S0140525X00076743

Harmon-Jones, E., Gable, P. A., & Peterson, C. K. (2010). The role of asymmetric frontal cortical activity in emotion-related phenomena: A review and update. *Biological Psychology*, *84*(3), 451–462. https://doi.org/10.1016/j.biopsycho.2009.08.010

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. https://doi.org/10.1111/j.0963-7214.2005.00337.x

Hartwright, C. E., Apperly, I. A., & Hansen, P. C. (2015). The special case of self-perspective inhibition in mental, but not non-mental, representation. *Neuropsychologia*, *67*, 183–192. https://doi.org/10.1016/j.neuropsychologia.2014.12.015

Heil, L., Colizoli, O., Hartstra, E., Kwisthout, J., Van Pelt, S., Van Rooij, I., & Bekkering, H. (2019). Processing of prediction errors in mentalizing areas. *Journal of Cognitive Neuroscience*, *31*(6), 900–912. https://doi.org/10.1162/jocn_a_01381

Helming, K. A., Strickland, B., & Jacob, P. (2014). Making sense of early false-belief understanding. *Trends in Cognitive Sciences*, *18*(4), 167–170. https://doi.org/10.1016/j.tics.2014.01.005

Hillebrandt, H., Friston, K. J., & Blakemore, S.-J. (2014). Effective connectivity during animacy perception – dynamic causal modelling of Human Connectome Project data. *Scientific Reports*, *4*(1), 6240. https://doi.org/10.1038/srep06240

Hirst, D. (1989). Relevance: Communication and cognition. *Mind & Language*, *4*(1–2), 138–146. https://doi.org/10.1111/j.1468-0017.1989.tb00246.x

Hoehl, S., & Wahl, S. (2012). Recording infant ERP data for cognitive research. *Developmental Neuropsychology*, *37*(3), 187–209. https://doi.org/10.1080/87565641.2011.627958

Hofmann, S. G., Doan, S. N., Sprung, M., Wilson, A., Ebesutani, C., Andrews, L. A., Curtiss, J., & Harris, P. L. (2016). Training children's theory-of-mind: A meta-analysis of controlled studies. *Cognition*, *150*, 200–212. https://doi.org/10.1016/j.cognition.2016.01.006

Hogrefe, G.-J., Wimmer, H., & Perner, J. (1986). Ignorance versus false belief: A developmental lag in attribution of epistemic states. *Child Development*, *57*(3), 567–582. https://doi.org/10.2307/1130337

Hopfinger, J. B., & Mangun, G. R. (2001). Tracking the influence of reflexive attention on sensory and cognitive processing. *Cognitive, Affective, & Behavioral Neuroscience*, *1*(1), 56–65. https://doi.org/10.3758/CABN.1.1.56

Hoptman, M. J., & Davidson, R. J. (1998). Baseline EEG asymmetries and performance on neuropsychological tasks. *Neuropsychologia*, *36*(12), 1343–1353. https://doi.org/10.1016/S0028-3932(98)00023-2

Howell, D. C. (2010). Multiple regression. In *Statistical Methods for psychology*. PWS-Kent Publishing Co.

Hughes, C. (1998). Executive function in preschoolers: Links with theory of mind and verbal ability. *British Journal of Developmental Psychology*, *16*(2), 233–253. https://doi.org/10.1111/j.2044-835X.1998.tb00921.x

Hughes, C., Cassidy, B. S., Faskowitz, J., Avena-Koenigsberger, A., Sporns, O., & Krendl, A. C. (2019). Age differences in specific neural connections within the default mode network underlie theory of mind. *NeuroImage*, *191*, 269–277. https://doi.org/10.1016/j.neuroimage.2019.02.024

Hyde, D. C., Aparicio Betancourt, M., & Simon, C. E. (2015). Human temporal-parietal junction spontaneously tracks others’ beliefs: A functional near-infrared spectroscopy study. *Human Brain Mapping*, *36*(12), 4831–4846. https://doi.org/10.1002/hbm.22953

Hyde, D. C., Simon, C. E., Ting, F., & Nikolaeva, J. I. (2018). Functional organization of the temporal-parietal junction for theory of mind in preverbal infants: A near-infrared spectroscopy study. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *38*(18), 4264–4274. https://doi.org/10.1523/JNEUROSCI.0264-17.2018

Iacoboni, M. (2005). Neural mechanisms of imitation. *Current Opinion in Neurobiology*, *15*(6), 632–637. https://doi.org/10.1016/j.conb.2005.10.010

IBM Corp. (2023). *IBM SPSS Statistics for Windows, Version 29.0.2.0* (Version 29.0.2.0) [Computer software]. IBM Corp.

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*. https://doi.org/10.3389/fpsyg.2016.01501

Jones, N. A., Field, T., Davalos, M., & Pickens, J. (1997). EEG stability in infants/children of depressed mothers. *Child Psychiatry and Human Development*, *28*(2), 59–70. https://doi.org/10.1023/A:1025197101496

Jouravlev, O., Schwartz, R., Ayyash, D., Mineroff, Z., Gibson, E., & Fedorenko, E. (2019). Tracking colisteners’ knowledge states during language comprehension. *Psychological Science*, *30*(1), 3–19. https://doi.org/10.1177/0956797618807674

Kail, R. (1991). Developmental change in speed of processing during childhood and adolescence. *Psychological Bulletin*, *109*(3), 490–501. https://doi.org/10.1037/0033-2909.109.3.490

Kaletsch, K., & Liszkowski, U. (2024). A new online paradigm to measure spontaneous pointing in infants and caregivers. *Infant Behavior and Development*, *74*, 101907. https://doi.org/10.1016/j.infbeh.2023.101907

Kaltefleiter, L. J. (2022). *Children’s Mental State Reasoning Abilities in the Third Year of Life and Relations with Complement Syntax, Mental State Vocabulary, and Perspective-Shifting Discourse* [PhD Thesis]. Ludwig-Maximilians-Universität München.

Kaltefleiter, L. J., Schuwerk, T., Wiesmann, C. G., Kristen-Antonow, S., Jarvers, I., & Sodian, B. (2022). Evidence for goal- and mixed evidence for false belief-based action prediction in 2- to 4-year-old children: A large-scale longitudinal anticipatory looking replication study. *Developmental Science*, *25*(4). https://doi.org/10.1111/desc.13224

Kaltefleiter, L. J., Sodian, B., Kristen-Antonow, S., Grosse Wiesmann, C., & Schuwerk, T. (2021). Does syntax play a role in theory of mind development before the age of 3 years? *Infant Behavior and Development*, *64*, 101575. https://doi.org/10.1016/j.infbeh.2021.101575

Kampis, D., Askitis, D., & Southgate, V. (2024). *Altercentric bias in infants' encoding of object kind*.

Kampis, D., Karman, P., Csibra, G., Southgate, V., & Hernik, M. (2021). A two-lab direct replication attempt of Southgate, Senju and Csibra (2007). *Royal Society Open Science*, *8*(8), 210190.

Karmiloff-Smith, A. (1997). Crucial differences between developmental cognitive neuroscience and adult neuropsychology. *Developmental Neuropsychology*, *13*(4), 513–524. https://doi.org/10.1080/87565649709540693

Keil, A. (2013). Electro-and magnetoencephalography in the study of emotion. *The Cambridge Handbook of Human Affective Neuroscience*, *107*, 137–132.

Kikuno, H., Mitchell, P., & Ziegler, F. (2007). How do young children process beliefs about beliefs?: Evidence from response latency. *Mind & Language*, *22*(3), 297–316. https://doi.org/10.1111/j.1468-0017.2007.00310.x

Kim, K. J., & Bell, M. A. (2006). Frontal EEG asymmetry and regulation during childhood. *Annals of the New York Academy of Sciences*, *1094*(1), 308–312. https://doi.org/10.1196/annals.1376.040

Király, I., Oláh, K., Csibra, G., & Kovács, Á. M. (2018). Retrospective attribution of false beliefs in 3-year-old children. *Proceedings of the National Academy of Sciences*, *115*(45), 11477–11482. https://doi.org/10.1073/pnas.1803505115

Kliemann, D., Young, L., Scholz, J., & Saxe, R. (2008). The influence of prior record on moral judgment. *Neuropsychologia*, *46*(12), 2949–2957. https://doi.org/10.1016/j.neuropsychologia.2008.06.010

Klimesch, W. (1999). EEG alpha and theta oscillations reflect cognitive and memory performance: A review and analysis. *Brain Research Reviews*, *29*(2), 169–195. https://doi.org/10.1016/S0165-0173(98)00056-3

Kloo, D., Kaltefleiter, L. J., & Sodian, B. (2024). Early perspective taking predicts later cognitive flexibility: A longitudinal study. *Infant and Child Development*, e2537. https://doi.org/10.1002/icd.2537

Kloo, D., Kristen-Antonow, S., & Sodian, B. (2020). Progressing from an implicit to an explicit false belief understanding: A matter of executive control? *International Journal of Behavioral Development*, *44*(2), 107–115. https://doi.org/10.1177/0165025419850901

Kloo, D., Osterhaus, C., Kristen-Antonow, S., & Sodian, B. (2022). The impact of theory of mind and executive function on math and reading abilities: A longitudinal study. *Infant and Child Development*, *31*(6), e2356. https://doi.org/10.1002/icd.2356

Kloo, D., & Sodian, B. (2017). The developmental stability of inhibition from 2 to 5 years. *British Journal of Developmental Psychology*, *35*(4), 582–595. https://doi.org/10.1111/bjdp.12197

Kloo, D., Sodian, B., Kristen-Antonow, S., Kim, S., & Paulus, M. (2021). Knowing minds: Linking early perspective taking and later metacognitive insight. *British Journal of Developmental Psychology*, *39*(1), 39–53. https://doi.org/10.1111/bjdp.12359

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. (2007a). Children's and adults' neural bases of verbal and nonverbal 'theory of mind.' *Neuropsychologia*, *45*(7), 1522–1532. https://doi.org/10.1016/j.neuropsychologia.2006.11.017

Kobayashi, C., Glover, G. H., & Temple, E. (2007b). Cultural and linguistic effects on neural bases of 'theory of mind' in American and Japanese children. *Brain Research*, *1164*, 95–107. https://doi.org/10.1016/j.brainres.2007.06.022

Koster-Hale, J., Richardson, H., Velez, N., Asaba, M., Young, L., & Saxe, R. (2017). Mentalizing regions represent distributed, continuous, and abstract dimensions of others' beliefs. *NeuroImage*, *161*, 9–18. https://doi.org/10.1016/j.neuroimage.2017.08.026

Kovács, Á. M., Téglás, E., & Endress, A. D. (2010). The social sense: Susceptibility to others' beliefs in human infants and adults. *Science*, *330*(6012), 1830–1834. https://doi.org/10.1126/science.1190792

Krall, S. C., Rottschy, C., Oberwelland, E., Bzdok, D., Fox, P. T., Eickhoff, S. B., Fink, G. R., & Konrad, K. (2015). The role of the right temporoparietal junction in attention and social interaction as revealed by ALE meta-analysis. *Brain Structure and Function*, *220*(2), 587–604. https://doi.org/10.1007/s00429-014-0803-z

Krause, L., Enticott, P. G., Zangen, A., & Fitzgerald, P. B. (2012). The role of medial prefrontal cortex in theory of mind: A deep rTMS study. *Behavioural Brain Research*, *228*(1), 87–90. https://doi.org/10.1016/j.bbr.2011.11.037

Kühn-Popp, N., Kristen, S., Paulus, M., Meinhardt, J., & Sodian, B. (2016). Left hemisphere EEG coherence in infancy predicts infant declarative pointing and preschool epistemic language. *Social Neuroscience*, *11*(1), 49–59. https://doi.org/10.1080/17470919.2015.1024887

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. https://doi.org/10.1016/j.neuroscience.2013.06.017

Kulke, L., Johannsen, J., & Rakoczy, H. (2019). Why can some implicit theory of mind tasks be replicated and others cannot? A test of mentalizing versus submentalizing accounts. *PloS One*, *14*(3), e0213772.

Kulke, L., & Rakoczy, H. (2018). Implicit Theory of Mind – An overview of current replications and non-replications. *Data in Brief*, *16*, 101–104. https://doi.org/10.1016/j.dib.2017.11.016

Kulke, L., & Rakoczy, H. (2019). Testing the role of verbal narration in implicit theory of mind tasks. *Journal of Cognition and Development*, *20*(1), 1–14.

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. https://doi.org/10.1016/j.cogdev.2017.09.001

Kulke, L., von Duhn, B., Schneider, D., & Rakoczy, H. (2018). Is implicit theory of mind a real and robust phenomenon? Results from a systematic replication study. *Psychological Science*, *29*(6), 888–900.

Kulke, L., Wübker, M., & Rakoczy, H. (2019). Is implicit theory of mind real but hard to detect? Testing adults with different stimulus materials. *Royal Society Open Science*, *6*(7), 190068.

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*(1), 621–647. https://doi.org/10.1146/annurev.psych.093008.131123

Lau, E., Phillips, C., & Poeppel, D. (2009). A cortical network for semantics: (De)constructing the N400. *Nature Reviews. Neuroscience*, *9*, 920–933. https://doi.org/10.1038/nrn2532

Le Petit, M., Eustache, F., Perrier, J., de La Sayette, V., Desgranges, B., & Laisney, M. (2022). Functional connectivity of the medial prefrontal cortex related to mindreading abilities. *Cerebral Cortex Communications*, *3*, 1–13. https://doi.org/10.1093/texcom/tgac032

Lee, J., Quintana, J., Nori, P., & Green, M. F. (2011). Theory of mind in schizophrenia: Exploring neural mechanisms of belief attribution. *Social Neuroscience*, *6*(5–6), 569–581. https://doi.org/10.1080/17470919.2011.620774

Lee, S. M., & McCarthy, G. (2016). Functional heterogeneity and convergence in the right temporoparietal junction. *Cerebral Cortex*, *26*(3), 1108–1116. https://doi.org/10.1093/cercor/bhu292

Leekam, S. R., & Perner, J. (1991). Does the autistic child have a metarepresentational deficit? *Cognition*, *40*(3), 203–218. https://doi.org/10.1016/0010-0277(91)90025-Y

Leslie, A. M. (1987). Pretense and representation: The origins of “theory of mind.” *Psychological Review*, *94*, 412–426.

Leslie, A. M. (1988). Some implications of pretense for mechanisms underlying the child’s theory of mind. In J. W. Astington, P. L. Harris, & D. R. Olson (Eds.), *Developing theories of mind* (pp. 19–46). Cambridge University Press.

Leslie, A. M. (1994a). Pretending and believing: Issues in the theory of ToMM. *Cognition*, *50*(1–3), 211–238. https://doi.org/10.1016/0010-0277(94)90029-9

Leslie, A. M. (1994b). ToMM, ToBY, and Agency: Core architecture and domain specificity. In L. A. Hirschfeld & S. A. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 119–148). Cambridge University Press.

Leslie, A. M., Friedman, O., & German, T. P. (2004). Core mechanisms in ‘theory of mind.’ *Trends in Cognitive Sciences*, *8*(12), 528–533. https://doi.org/10.1016/j.tics.2004.10.001

Leslie, A. M., & Polizzi, P. (1998). Inhibitory processing in the false belief task: Two conjectures. *Developmental Science*, *1*(2), 247–253. https://doi.org/10.1111/1467-7687.00038

Leslie, A. M., & Thaiss, L. (1992). Domain specificity in conceptual development: Neuropsychological evidence from autism. *Cognition*, *43*(3), 225–251. https://doi.org/10.1016/0010-0277(92)90013-8

Li, G., Nie, J., Wang, L., Shi, F., Lyall, A. E., Lin, W., Gilmore, J. H., & Shen, D. (2014). Mapping longitudinal hemispheric structural asymmetries of the human cerebral cortex from birth to 2 years of age. *Cerebral Cortex*, *24*(5), 1289–1300. https://doi.org/10.1093/cercor/bhs413

Li, S., Meinhardt, J., & Sodian, B. (2023). Neural correlates of theory of mind development in young children. *International Journal on Neuropsychology and Behavioural Sciences*, *4*(1), 1–2.

Licata, M., Paulus, M., Kühn-Popp, N., Meinhardt, J., & Sodian, B. (2015). Infant frontal asymmetry predicts child emotional availability. *International Journal of Behavioral Development*, *39*(6), 492–496. https://doi.org/10.1177/0165025415576816

Liu, D., Meltzoff, A. N., & Wellman, H. M. (2009a). Neural correlates of belief- and desire-reasoning. *Child Development*, *80*(4), 1163–1171. https://doi.org/10.1111/j.1467-8624.2009.01323.x

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. https://doi.org/10.1097/01.wnr.0000123388.87650.06

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. https://doi.org/10.1111/j.1467-8624.2009.01262.x

Liu, D., Wellman, H., Tardif, T., & Sabbagh, M. (2008). Theory of mind development in Chinese children: A meta-analysis of false-belief understanding across cultures and languages. *Developmental Psychology*, *44*, 523–531. https://doi.org/10.1037/0012-1649.44.2.523

Low, J. (2010). Preschoolers’ implicit and explicit false belief understanding: Relations with complex syntactical mastery. *Child Development*, *81*(2), 597–615. https://doi.org/10.1111/j.1467-8624.2009.01418.x

Low, J., Apperly, I. A., Butterfill, S. A., & Rakoczy, H. (2016). Cognitive architecture of belief reasoning in children and adults: A primer on the two-systems account. *Child Development Perspectives*, *10*(3), 184–189.

Low, J., & Watts, J. (2013). Attributing false beliefs about object identity reveals a signature blind spot in humans’ efficient mind-reading system. *Psychological Science*, *24*(3), 305–311.

MacLean, M. H., Arnell, K. M., & Cote, K. A. (2012). Resting EEG in alpha and beta bands predicts individual differences in attentional blink magnitude. *Brain and Cognition*, *78*(3), 218–229. https://doi.org/10.1016/j.bandc.2011.12.010

Mahy, C. E. V., Moses, L. J., & Pfeifer, J. H. (2014). How and where: Theory-of-mind in the brain. *Developmental Cognitive Neuroscience*, *9*, 68–81. https://doi.org/10.1016/j.dcn.2014.01.002

Marcovitch, S., O’Brien, M., Calkins, S. D., Leerkes, E. M., Weaver, J. M., & Levine, D. W. (2015). A longitudinal assessment of the relation between executive function and theory of mind at 3, 4, and 5 years. *Cognitive Development*, *33*, 40–55. https://doi.org/10.1016/j.cogdev.2014.07.001

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

Mars, R. B., Neubert, F.-X., Noonan, M. P., Sallet, J., Toni, I., & Rushworth, M. F. S. (2012a). On the relationship between the “default mode network” and the “social brain.” *Frontiers in Human Neuroscience*, *6*. https://doi.org/10.3389/fnhum.2012.00189

Mars, R. B., Sallet, J., Schüffelgen, U., Jbabdi, S., Toni, I., & Rushworth, M. F. S. (2012b). Connectivity-based subdivisions of the human right “temporoparietal junction area”: Evidence for different areas participating in different cortical networks. *Cerebral Cortex*, *22*(8), 1894–1903. https://doi.org/10.1093/cercor/bhr268

Marshall, P. J., Bar-Haim, Y., & Fox, N. A. (2002). Development of the EEG from 5 months to 4 years of age. *Clinical Neurophysiology*, *113*(8), 1199–1208. https://doi.org/10.1016/S1388-2457(02)00163-3

Martin, A., & Santos, L. R. (2016). What cognitive representations support primate theory of mind? *Trends in Cognitive Sciences*, *20*(5), 375–382. https://doi.org/10.1016/j.tics.2016.03.005

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. https://doi.org/10.1016/j.neuroimage.2012.07.019

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. https://doi.org/10.1016/j.dcn.2010.08.001

Melchers, P., & Melchers, M. (2015). *Kaufman assessment battery for children, second edition (KABC-II). (Deutschsprachige Fassung)*. NCS Pearson Inc.

Meltzoff, A. N., & Decety, J. (2003). What imitation tells us about social cognition: A rapprochement between developmental psychology and cognitive neuroscience. *Philosophical Transactions of the Royal*

Society of London. Series B: Biological Sciences, *358*(1431), 491–500. https://doi.org/10.1098/rstb.2002.1261

Meristo, M., Morgan, G., Geraci, A., Iozzi, L., Hjelmquist, E., Surian, L., & Siegal, M. (2012). Belief attribution in deaf and hearing infants. *Developmental Science*, *15*(5), 633–640. https://doi.org/10.1111/j.1467-7687.2012.01155.x

Meristo, M., Strid, K., & Hjelmquist, E. (2016). Early conversational environment enables spontaneous belief attribution in deaf children. *Cognition*, *157*, 139–145. https://doi.org/10.1016/j.cognition.2016.08.023

Millett, D. (2001). Hans berger: From psychic energy to the EEG. *Perspectives in Biology and Medicine*, *44*(4), 522–542. https://doi.org/10.1353/pbm.2001.0070

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. https://doi.org/10.1111/j.1467-8624.2007.01018.x

Mitchell, J. P. (2008). Activity in right temporo-parietal junction is not selective for theory-of-mind. *Cerebral Cortex*, *18*(2), 262–271. https://doi.org/10.1093/cercor/bhm051

Mitchell, P., & Lacohée, H. (1991). Children's early understanding of false belief. *Cognition*, *39*(2), 107–127. https://doi.org/10.1016/0010-0277(91)90040-B

Moessnang, C., Otto, K., Bilek, E., Schäfer, A., Baumeister, S., Hohmann, S., Poustka, L., Brandeis, D., Banaschewski, T., Tost, H., & Meyer-Lindenberg, A. (2017). Differential responses of the dorsomedial prefrontal cortex and right posterior superior temporal sulcus to spontaneous mentalizing. *Human Brain Mapping*, *38*(8), 3791–3803. https://doi.org/10.1002/hbm.23626

Molenberghs, P., Johnson, H., Henry, J. D., & Mattingley, J. B. (2016). Understanding the minds of others: A neuroimaging meta-analysis. *Neuroscience & Biobehavioral Reviews*, *65*, 276–291. https://doi.org/10.1016/j.neubiorev.2016.03.020

Moll, H., Khalulyan, A., & Moffett, L. (2017). 2.5-year-olds express suspense when others approach reality with false expectations. *Child Development*, *88*(1), 114–122.

Moll, H., & Meltzoff, A. N. (2012). Joint attention as the fundamental basis of understanding perspectives. In A. Seemann (Ed.), *Joint Attention* (pp. 393–414). The MIT Press. https://doi.org/10.7551/mitpress/8841.003.0019

Moraczewski, D., Chen, G., & Redcay, E. (2018). Inter-subject synchrony as an index of functional specialization in early childhood. *Scientific Reports*, *8*(1), 2252. https://doi.org/10.1038/s41598-018-20600-0

Moses, L. J. (2001). Executive accounts of theory-of-mind development. *Child Development*, *72*(3), 688–690. https://doi.org/10.1111/1467-8624.00306

Müller, B. C. N., Kühn-Popp, N., Meinhardt, J., Sodian, B., & Paulus, M. (2015). Long-term stability in children's frontal EEG alpha asymmetry between 14-months and 83-months. *International Journal of Developmental Neuroscience*, *41*(1), 110–114. https://doi.org/10.1016/j.ijdevneu.2015.01.002

Müller, B. C. N., Marx, A. K. G., Paulus, M., & Meinhardt, J. (2018). Frontal electroencephalogram alpha asymmetry relates to implicit achievement motives: A pilot study. *Mind, Brain, and Education*, *12*(2), 82–89. https://doi.org/10.1111/mbe.12175

Mundy, P., Card, J., & Fox, N. (2000). EEG correlates of the development of infant joint attention skills. *Developmental Psychobiology*, *36*(4), 325–338. https://doi.org/10.1002/(SICI)1098-2302(200005)36:4<325::AID-DEV7>3.0.CO;2-F

Naughtin, C. K., Horne, K., Schneider, D., Venini, D., York, A., & Dux, P. E. (2017). Do implicit and explicit belief processing share neural substrates? *Human Brain Mapping*, *38*(9), 4760–4772. https://doi.org/10.1002/hbm.23700

Nelson, C. A., & McCleery, J. P. (2008). Use of event-related potentials in the study of typical and atypical development. *Journal of the American Academy of Child & Adolescent Psychiatry*, *47*(11), 1252–1261. https://doi.org/10.1097/CHI.0b013e318185a6d8

Oliver, L. D., Vieira, J. B., Neufeld, R. W. J., Dziobek, I., & Mitchell, D. G. V. (2018). Greater involvement of action simulation mechanisms in emotional vs cognitive empathy. *Social Cognitive and Affective Neuroscience*, *13*(4), 367–380. https://doi.org/10.1093/scan/nsy013

Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science*, *308*(5719), 255–258. https://doi.org/10.1126/science.1107621

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2010). Fieldtrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, *2011*, e156869. https://doi.org/10.1155/2011/156869

Osterhaus, C., Kristen-Antonow, S., Kloo, D., & Sodian, B. (2022). Advanced scaling and modeling of children's theory of mind competencies: Longitudinal findings in 4- to 6-year-olds. *International Journal of Behavioral Development*, 016502542210773. https://doi.org/10.1177/01650254221077334

Özdem, C., Brass, M., Van Der Cruyssen, L., & Van Overwalle, F. (2017). The overlap between false belief and spatial reorientation in the temporo-parietal junction: The role of input modality and task. *Social Neuroscience*, *12*(2), 207–217. https://doi.org/10.1080/17470919.2016.1143027

Papo, D. (2013). Why should cognitive neuroscientists study the brain's resting state? *Frontiers in Human Neuroscience*, *7*. https://doi.org/10.3389/fnhum.2013.00045

Paulus, M., Hunnius, S., & Bekkering, H. (2013). Neurocognitive mechanisms underlying social learning in infancy: Infants' neural processing of the effects of others' actions. *Social Cognitive and Affective Neuroscience*, *8*(7), 774–779. https://doi.org/10.1093/scan/nss065

Paulus, M., Kühn-Popp, N., Licata, M., Sodian, B., & Meinhardt, J. (2013). Neural correlates of prosocial behavior in infancy: Different neurophysiological mechanisms support the emergence of helping and comforting. *NeuroImage*, *66*, 522–530. https://doi.org/10.1016/j.neuroimage.2012.10.041

Perner, J. (1988). Developing semantics for theories of mind: From propositional attitudes to mental representation. In J. W. Astington, P. L. Harris, & D. R. Olson (Eds.), *Developing theories of mind* (pp. 141–172). Cambridge University Press.

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

Perner, J., Aichhorn, M., Kronbichler, M., Staffen, W., & Ladurner, G. (2006). Thinking of mental and other representations: The roles of left and right temporo-parietal junction. *Social Neuroscience*, *1*(3–4), 245–258. https://doi.org/10.1080/17470910600989896

Perner, J., Baker, S., & Hutton, D. (1994). Prelief: The conceptual origins of belief and pretence. In C. Lewis & P. Mitchell (Eds.), *Children's early understanding of mind: Origins and development* (pp. 261–286).

Perner, J., Huemer, M., & Leahy, B. (2015). Mental files and belief: A cognitive theory of how children represent belief and its intensionality. *Cognition*, *145*, 77–88. https://doi.org/10.1016/j.cognition.2015.08.006

Perner, J., & Lang, B. (1999). Development of theory of mind and executive control. *Trends in Cognitive Sciences*, *3*(9), 337–344. https://doi.org/10.1016/S1364-6613(99)01362-5

Perner, J., & Leahy, B. (2016). Mental files in development: Dual naming, false belief, identity and intensionality. *Review of Philosophy and Psychology*, *7*(2), 491–508. https://doi.org/10.1007/s13164-015-0235-6

Perner, J., & Leekam, S. (2008). The curious incident of the photo that was accused of being false: Issues of domain specificity in development, autism, and brain imaging. *Quarterly Journal of Experimental Psychology*, *61*(1), 76–89. https://doi.org/10.1080/17470210701508756

Perner, J., Leekam, S. R., & Wimmer, H. (1987). Three-year-olds' difficulty with false belief: The case for a conceptual deficit. *British Journal of Developmental Psychology*, *5*(2), 125–137. https://doi.org/10.1111/j.2044-835X.1987.tb01048.x

Perner, J., Mauer, M. C., & Hildenbrand, M. (2011). Identity: Key to children's understanding of belief. *Science*, *333*(6041), 474–477. https://doi.org/10.1126/science.1201216

Perner, J., Rendl, B., & Garnham, A. (2007). Objects of desire, thought, and reality: Problems of anchoring discourse referents in development.

Mind & Language, *22*(5), 475–513. https://doi.org/10.1111/j.1468-0017.2007.00317.x

Perner, J., & Roessler, J. (2012). From infants’ to children’s appreciation of belief. *Trends in Cognitive Sciences*, *16*(10), 519–525. https://doi.org/10.1016/j.tics.2012.08.004

Peterson, C. C., Wellman, H. M., & Liu, D. (2005). Steps in theory-of-mind development for children with deafness or autism. *Child Development*, *76*(2), 502–517. https://doi.org/10.1111/j.1467-8624.2005.00859.x

Picton, T. W., Bentin, S., Berg, P., Donchin, E., Hillyard, S. A., Johnson, R., Miller, G. A., Ritter, W., Ruchkin, D. S., Rugg, M. D., & Taylor, M. J. (2000). Guidelines for using human event-related potentials to study cognition: Recording standards and publication criteria. *Psychophysiology*, *37*(2), 127–152. https://doi.org/10.1111/1469-8986.3720127

Pitchford, B., & Arnell, K. M. (2019). Resting EEG in alpha and beta bands predicts individual differences in attentional breadth. *Consciousness and Cognition*, *75*, 102803. https://doi.org/10.1016/j.concog.2019.102803

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology*, *118*(10), 2128–2148. https://doi.org/10.1016/j.clinph.2007.04.019

Polich, J. (2012). Neuropsychology of P300. In E. S. Kappenman & S. J. Luck (Eds.), *The Oxford handbook of event-related potential components* (pp. 159–188).

Poulin-Dubois, D. (2020). Chapter 7 - theory of mind development: State of the science and future directions. In S. Hunnius & M. Meyer (Eds.), *Progress in Brain Research* (Vol. 254, pp. 141–166). Elsevier. https://doi.org/10.1016/bs.pbr.2020.05.021

Poulin-Dubois, D., Azar, N., Elkaim, B., & Burnside, K. (2020). Testing the stability of theory of mind: A longitudinal approach. *PLOS ONE*, *15*(11), e0241721. https://doi.org/10.1371/journal.pone.0241721

Poulin-Dubois, D., Goldman, E. J., Meltzer, A., & Psaradellis, E. (2023). Discontinuity from implicit to explicit theory of mind from infancy to preschool age. *Cognitive Development*, *65*, 101273. https://doi.org/10.1016/j.cogdev.2022.101273

Poulin-Dubois, D., Polonia, A., & Yott, J. (2013). Is false belief skin-deep? The agent's eye status influences infants' reasoning in belief-inducing situations. *Journal of Cognition and Development*, *14*(1), 87–99. https://doi.org/10.1080/15248372.2011.608198

Poulin-Dubois, D., & Yott, J. (2018). Probing the depth of infants' theory of mind: Disunity in performance across paradigms. *Developmental Science*, *21*(4), e12600.

Powell, L. J., Hobbs, K., Bardis, A., Carey, S., & Saxe, R. (2018). Replications of implicit theory of mind tasks with varying representational demands. *Cognitive Development*, *46*, 40–50.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, *1*(4), 515–526. https://doi.org/10.1017/S0140525X00076512

Pyers, J. E., & Senghas, A. (2009). Language promotes false-belief understanding: Evidence from learners of a new sign language. *Psychological Science*, *20*(7), 805–812. https://doi.org/10.1111/j.1467-9280.2009.02377.x

Pylyshyn, Z. W. (1978). When is attribution of beliefs justified? *Behavioral and Brain Sciences*, *1*(4), 592–593. https://doi.org/10.1017/S0140525X00076895

Quesque, F., Apperly, I., Baillargeon, R., Baron-Cohen, S., Becchio, C., Bekkering, H., Bernstein, D., Bertoux, M., Bird, G., Bukowski, H., Burgmer, P., Carruthers, P., Catmur, C., Dziobek, I., Epley, N., Erle, T. M., Frith, C., Frith, U., Galang, C. M., … Brass, M. (2024). Defining key concepts for mental state attribution. *Communications Psychology*, *2*(1), 29. https://doi.org/10.1038/s44271-024-00077-6

Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., & Shulman, G. L. (2001). A default mode of brain function. *Proceedings of the National Academy of Sciences*, *98*(2), 676–682. https://doi.org/10.1073/pnas.98.2.676

Rakoczy, H. (2022). Foundations of theory of mind and its development in early childhood. *Nature Reviews Psychology*, *1*(4), Article 4. https://doi.org/10.1038/s44159-022-00037-z

Reznik, S. J., & Allen, J. J. B. (2018). Frontal asymmetry as a mediator and moderator of emotion: An updated review. *Psychophysiology*, *55*(1), e12965. https://doi.org/10.1111/psyp.12965

Richardson, H., Lisandrelli, G., Riobueno-Naylor, A., & Saxe, R. (2018). Development of the social brain from age three to twelve years. *Nature Communications*, *9*(1), 1027. https://doi.org/10.1038/s41467-018-03399-2

Richardson, H., & Saxe, R. (2020a). Chapter 21—Early signatures of and developmental change in brain regions for theory of mind. In J. Rubenstein, P. Rakic, B. Chen, & K. Y. Kwan (Eds.), *Neural Circuit and Cognitive Development (Second Edition)* (pp. 467–484). Academic Press. https://doi.org/10.1016/B978-0-12-814411-4.00021-4

Richardson, H., & Saxe, R. (2020b). Development of predictive responses in theory of mind brain regions. *Developmental Science*, *23*(1). https://doi.org/10.1111/desc.12863

Riggins, T., & Scott, L. S. (2020). P300 development from infancy to adolescence. *Psychophysiology*, *57*(7). https://doi.org/10.1111/psyp.13346

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, *27*(Volume 27, 2004), 169–192. https://doi.org/10.1146/annurev.neuro.27.070203.144230

Robinson, E. J., & Mitchell, P. (1995). Masking of children’s early understanding of the representational mind: Backwards explanation versus prediction. *Child Development*, *66*(4), 1022–1039. JSTOR. https://doi.org/10.2307/1131796

Rochat, P., & Striano, T. (2002). Who’s in the mirror? Self–other discrimination in specular images by four- and nine-month-old infants. *Child Development*, *73*(1), 35–46. https://doi.org/10.1111/1467-8624.00390

Rösler, F., Heil, M., & Röder, B. (1997). Slow negative brain potentials as reflections of specific modular resources of cognition. *Biological*

Psychology, *45*(1), 109–141. https://doi.org/10.1016/S0301-0511(96)05225-8

Roth, D., & Leslie, A. M. (1998). Solving belief problems: Toward a task analysis. *Cognition*, *66*(1), 1–31. https://doi.org/10.1016/S0010-0277(98)00005-5

Rothmayr, C., Sodian, B., Hajak, G., Döhnel, K., Meinhardt, J., & Sommer, M. (2011). Common and distinct neural networks for false-belief reasoning and inhibitory control. *NeuroImage*, *56*(3), 1705–1713. https://doi.org/10.1016/j.neuroimage.2010.12.052

Rowe, A. D., Bullock, P. R., Polkey, C. E., & Morris, R. G. (2001). “Theory of mind” impairments and their relationship to executive functioning following frontal lobe excisions. *Brain*, *124*(3), 600–616. https://doi.org/10.1093/brain/124.3.600

RStudio Team. (2024). [Computer software] (Version 2024.09.0+375). http://www.rstudio.com/

Ruchkin, D. S., Canoune, H. L., Johnson, R., & Ritter, W. (1995). Working memory and preparation elicit different patterns of slow wave event-related brain potentials. *Psychophysiology*, *32*(4), 399–410. https://doi.org/10.1111/j.1469-8986.1995.tb01223.x

Rueckert, L., & Naybar, N. (2008). Gender differences in empathy: The role of the right hemisphere. *Brain and Cognition*, *67*(2), 162–167. https://doi.org/10.1016/j.bandc.2008.01.002

Rueschemeyer, S.-A., Gardner, T., & Stoner, C. (2015). The social N400 effect: How the presence of other listeners affects language comprehension. *Psychonomic Bulletin & Review*, *22*(1), 128–134. https://doi.org/10.3758/s13423-014-0654-x

Ruffman, T. (2014). To belief or not belief: Children’s theory of mind. *Developmental Review*, *34*(3), 265–293. https://doi.org/10.1016/j.dr.2014.04.001

Ruffman, T., Taumoepeau, M., & Perkins, C. (2012). Statistical learning as a basis for social understanding in children. *British Journal of Developmental Psychology*, *30*(1), 87–104. https://doi.org/10.1111/j.2044-835X.2011.02045.x

Sabbagh, M. A. (2013). Brain electrophysiological studies of theory of mind. In S. Baron-Cohen, H. Tager–Flusberg, & M. V. Lombardo (Eds.), *Understanding Other Minds: Perspectives from Developmental Social Neuroscience*. Oxford University Press.

Sabbagh, M. A., Bowman, L. C., Evraire, L. E., & Ito, J. M. B. (2009). Neurodevelopmental correlates of theory of mind in preschool children. *Child Development*, *80*(4), 1147–1162. https://doi.org/10.1111/j.1467-8624.2009.01322.x

Sabbagh, M. A., & Flynn, J. (2006). Mid-frontal EEG alpha asymmetries predict individual differences in one aspect of theory of mind: Mental state decoding. *Social Neuroscience*, *1*(3–4), 299–308. https://doi.org/10.1080/17470910601029163

Sabbagh, M. A., & Taylor, M. (2000). Neural correlates of theory-of-mind reasoning: An event-related potential study. *Psychological Science*, *11*(1), 46–50. https://doi.org/10.1111/1467-9280.00213

Sabbagh, M. A., Xu, F., Carlson, S. M., Moses, L. J., & Lee, K. (2006). The development of executive functioning and theory of mind: A comparison of Chinese and U.S. preschoolers. *Psychological Science*, *17*(1), 74–81. https://doi.org/10.1111/j.1467-9280.2005.01667.x

Saby, J. N., & Marshall, P. J. (2012). The utility of EEG band power analysis in the study of infancy and early childhood. *Developmental Neuropsychology*, *37*(3), 253–273. https://doi.org/10.1080/87565641.2011.614663

Sai, L., Shang, S., Tay, C., Liu, X., Sheng, T., Fu, G., Ding, X. P., & Lee, K. (2021). Theory of mind, executive function, and lying in children: A meta-analysis. *Developmental Science*, *24*(5), e13096. https://doi.org/10.1111/desc.13096

Samson, D., Apperly, I. A., Chiavarino, C., & Humphreys, G. W. (2004). Left temporoparietal junction is necessary for representing someone else's belief. *Nature Neuroscience*, *7*(5), Article 5. https://doi.org/10.1038/nn1223

Saracho, O. N. (2014). Theory of mind theoretical foundations. In O. N. Saracho (Ed.), *Contemporary perspectives on research in theory of mind in early childhood education*. Information Age Publishing.

Saxe, R. (2009). Theory of mind (neural basis). In W. P. Banks (Ed.), *Encyclopedia of Consciousness* (pp. 401–409). Elsevier. https://doi.org/10.1016/B978-012373873-8.00078-5

Saxe, R., Carey, S., & Kanwisher, N. (2004a). Understanding other minds: Linking developmental psychology and functional neuroimaging. *Annual Review of Psychology*, *55*(1), 87–124. https://doi.org/10.1146/annurev.psych.55.090902.142044

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. https://doi.org/10.1016/S1053-8119(03)00230-1

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. https://doi.org/10.1111/j.1467-9280.2006.01768.x

Saxe, R., Schulz, L. E., & Jiang, Y. V. (2006). Reading minds versus following rules: Dissociating theory of mind and executive control in the brain. *Social Neuroscience*, *1*(3–4), 284–298. https://doi.org/10.1080/17470910601000446

Saxe, R., & Wexler, A. (2005). Making sense of another mind: The role of the right temporo-parietal junction. *Neuropsychologia*, *43*(10), 1391–1399. https://doi.org/10.1016/j.neuropsychologia.2005.02.013

Saxe, R., Whitfield-Gabrieli, S., Scholz, J., & Pelphrey, K. A. (2009). Brain regions for perceiving and reasoning about other people in school-aged children. *Child Development*, *80*(4), 1197–1209. https://doi.org/10.1111/j.1467-8624.2009.01325.x

Saxe, R., Xiao, D.-K., Kovacs, G., Perrett, D. I., & Kanwisher, N. (2004b). A region of right posterior superior temporal sulcus responds to observed intentional actions. *Neuropsychologia*, *42*(11), 1435–1446. https://doi.org/10.1016/j.neuropsychologia.2004.04.015

Schilbach, L., Bzdok, D., Timmermans, B., Fox, P. T., Laird, A. R., Vogeley, K., & Eickhoff, S. B. (2012). Introspective minds: Using ALE meta-analyses to study commonalities in the neural correlates of emotional processing, social & unconstrained cognition. *PLoS ONE*, *7*(2), e30920. https://doi.org/10.1371/journal.pone.0030920

Schilbach, L., Eickhoff, S. B., Rotarska-Jagiela, A., Fink, G. R., & Vogeley, K. (2008). Minds at rest? Social cognition as the default mode of cognizing and its putative relationship to the “default system” of the brain. *Consciousness and Cognition*, *17*(2), 457–467. https://doi.org/10.1016/j.concog.2008.03.013

Schneider, D., Bayliss, A. P., Becker, S. I., & Dux, P. E. (2012). Eye movements reveal sustained implicit processing of others’ mental states. *Journal of Experimental Psychology: General*, *141*(3), 433–438. https://doi.org/10.1037/a0025458

Schneider, D., Slaughter, V. P., & Dux, P. E. (2017). Current evidence for automatic theory of mind processing in adults. *Cognition*, *162*, 27–31. https://doi.org/10.1016/j.cognition.2017.01.018

Scholl, B. J., & Leslie, A. M. (1999). Modularity, development and ‘theory of mind.’ *Mind & Language*, *14*(1), 131–153. https://doi.org/10.1111/1468-0017.00106

Scholl, B. J., & Leslie, A. M. (2001). Minds, modules, and meta-analysis. *Child Development*, *72*(3), 696–701. https://doi.org/10.1111/1467-8624.00308

Scholz, J., Triantafyllou, C., Whitfield-Gabrieli, S., Brown, E. N., & Saxe, R. (2009). Distinct regions of right temporo-parietal junction are selective for theory of mind and exogenous attention. *PLoS ONE*, *4*(3), e4869. https://doi.org/10.1371/journal.pone.0004869

Schüler, C., Berger, P., & Wiesmann, C. G. (2024a). *A dorsal versus ventral network for understanding others in the developing brain* (p. 2024.05.23.595476). bioRxiv. https://doi.org/10.1101/2024.05.23.595476

Schüler, C., Berger, P., & Wiesmann, C. G. (2024b). *Modeling the connectome of joint attention in infancy predicts Theory of Mind in preschool-age* (p. 2024.05.22.595346). bioRxiv. https://doi.org/10.1101/2024.05.22.595346

Schurz, M., Radua, J., Aichhorn, M., Richlan, F., & Perner, J. (2014). Fractionating theory of mind: A meta-analysis of functional brain imaging studies. *Neuroscience & Biobehavioral Reviews*, *42*, 9–34. https://doi.org/10.1016/j.neubiorev.2014.01.009

Schurz, M., Radua, J., Tholen, M. G., Maliske, L., Margulies, D. S., Mars, R. B., Sallet, J., & Kanske, P. (2021). Toward a hierarchical model of social cognition: A neuroimaging meta-analysis and integrative review of empathy and theory of mind. *Psychological Bulletin*, *147*(3), 293–327. https://doi.org/10.1037/bul0000303

Schurz, M., Tholen, M. G., Perner, J., Mars, R. B., & Sallet, J. (2017). Specifying the brain anatomy underlying temporo-parietal junction activations for theory of mind: A review using probabilistic atlases from different imaging modalities. *Human Brain Mapping*, *38*(9), 4788–4805. https://doi.org/10.1002/hbm.23675

Schuwerk, T., Döhnel, K., Sodian, B., Keck, I. R., Rupprecht, R., & Sommer, M. (2014). Functional activity and effective connectivity of the posterior medial prefrontal cortex during processing of incongruent mental states: MPFC and Incongruent Mental States. *Human Brain Mapping*, *35*(7), 2950–2965. https://doi.org/10.1002/hbm.22377

Schuwerk, T., Grosso, S. S., & Taylor, P. C. J. (2021). The influence of TMS of the rTPJ on attentional control and mentalizing. *Neuropsychologia*, *162*, 108054. https://doi.org/10.1016/j.neuropsychologia.2021.108054

Schuwerk, T., Kampis, D., Baillargeon, R., Biro, S., Bohn, M., Byers-Heinlein, K., Dörrenberg, S., Fisher, C., Franchin, L., & Fulcher, T. (2024). *Action anticipation based on an agent's epistemic state in toddlers and adults [Registered Report](Version posted online February 13, 2024)*.

Schuwerk, T., Kampis, D., Baillargeon, R., Biro, S., Bohn, M., Byers-Heinlein, K., Dörrenberg, S., Fisher, C., Franchin, L., Fulcher, T., Garbisch, I., Geraci, A., Grosse Wiesmann, C., Hamlin, K., Haun, D., Hepach, R., Hunnius, S., Hyde, D. C., Karman, P., … Rakoczy, H. (2021). *Action anticipation based on an agent's epistemic state in toddlers and adults*. https://doi.org/10.31234/osf.io/x4jbm

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. https://doi.org/10.1098/rsos.172273

Schuwerk, T., Schurz, M., Müller, F., Rupprecht, R., & Sommer, M. (2017). The rTPJ's overarching cognitive function in networks for attention and theory of mind. *Social Cognitive and Affective Neuroscience*, *12*(1), 157–168. https://doi.org/10.1093/scan/nsw163

Scott, R. M. (2017). The developmental origins of false-belief understanding. *Current Directions in Psychological Science*, *26*(1), 68–74. https://doi.org/10.1177/0963721416673174

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. https://doi.org/10.1111/j.1467-8624.2009.01324.x

Scott, R. M., & Baillargeon, R. (2017). Early false-belief understanding. *Trends in Cognitive Sciences*, *21*(4), 237–249. https://doi.org/10.1016/j.tics.2017.01.012

Scott, R. M., 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.

Scott, R. M., Roby, E., & Baillargeon, R. (2022). How sophisticated is infants' theory of mind? In O. Houdé & G. Borst (Eds.), *The Cambridge Handbook of Cognitive Development* (1st ed., pp. 242–268). Cambridge University Press. https://doi.org/10.1017/9781108399838.015

Scott, R. M., Roby, E., & Setoh, P. (2020). 2.5-year-olds succeed in identity and location elicited-response false-belief tasks with adequate response practice. *Journal of Experimental Child Psychology*, *198*, 104890. https://doi.org/10.1016/j.jecp.2020.104890

Sebastian, C. L., Fontaine, N. M. G., Bird, G., Blakemore, S.-J., De Brito, S. A., McCrory, E. J. P., & Viding, E. (2012). Neural processing associated with cognitive and affective theory of mind in adolescents and adults. *Social Cognitive and Affective Neuroscience*, *7*(1), 53–63. https://doi.org/10.1093/scan/nsr023

Senju, A., Southgate, V., Snape, C., Leonard, M., & Csibra, G. (2011). Do 18-month-olds really attribute mental states to others?: A critical test. *Psychological Science*, *22*(7), 878–880. https://doi.org/10.1177/0956797611411584

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. https://doi.org/10.1073/pnas.1609203113

Shamay-Tsoory, S. G., Tomer, R., Berger, B. D., Goldsher, D., & Aharon-Peretz, J. (2005). Impaired “affective theory of mind” is associated with right ventromedial prefrontal damage. *Cognitive and Behavioral Neurology*, *18*(1), 55. https://doi.org/10.1097/01.wnn.0000152228.90129.99

Siegal, M., Carrington, J., & Radel, M. (1996). Theory of mind and pragmatic understanding following right hemisphere damage. *Brain and Language*, *53*(1), 40–50. https://doi.org/10.1006/brln.1996.0035

Sodian, B. (1991). The development of deception in young children. *British Journal of Developmental Psychology*, *9*(1), 173–188. https://doi.org/10.1111/j.2044-835X.1991.tb00869.x

Sodian, B. (2004). Theory of mind—The case for conceptual development. In W. Schneider, R. Schumann-Hengsteler, & B. Sodian (Eds.), *Young Children’s Cognitive Development*. Psychology Press.

Sodian, B. (2011). Theory of mind in infancy. *Child Development Perspectives*, *5*(1), 39–43. https://doi.org/10.1111/j.1750-8606.2010.00152.x

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. 301–320). Oxford University Press.

Sodian, B. (2023). Reply to Poulin-Dubois et al. (2023): Replication problems concerning both implicit and explicit false belief reasoning greatly reduced the chance of finding longitudinal correlations. *Cognitive Development*, *65*, 101294. https://doi.org/10.1016/j.cogdev.2022.101294

Sodian, B., & Frith, U. (1992). Deception and sabotage in autistic, retarded and normal children. *Journal of Child Psychology and Psychiatry*, *33*(3), 591–605. https://doi.org/10.1111/j.1469-7610.1992.tb00893.x

Sodian, B., & Hülsken, C. (2014). The developmental relation of theory of mind and executive functions: A study of advanced theory of mind abilities in children with attention deficit hyperactivity disorder. In *Young Children's Cognitive Development* (pp. 175–187). Psychology Press.

Sodian, B., Kaltefleiter, L. J., Schuwerk, T., & Kloo, D. (2024). Continuity in false belief understanding from 33 to 52 months of age. *Journal of Experimental Child Psychology*, *247*, 106039. https://doi.org/10.1016/j.jecp.2024.106039

Sodian, B., & Kristen, S. (2016). Chapter 5 theory of mind. In J. A. Greene, W. A. Sandoval, & I. Bråten (Eds.), *Handbook of Epistemic Cognition*. Routledge.

Sodian, B., & Kristen-Antonow, S. (2015). Declarative joint attention as a foundation of theory of mind. *Developmental Psychology*, *51*(9), 1190–1200. https://doi.org/10.1037/dev0000039

Sodian, B., Kristen-Antonow, S., & Kloo, D. (2020). How does children's theory of mind become explicit? A review of longitudinal findings. *Child Development Perspectives*, *14*(3), 171–177. https://doi.org/10.1111/cdep.12381

Sodian, B., Licata, M., Kristen-Antonow, S., Paulus, M., Killen, M., & Woodward, A. (2016). Understanding of goals, beliefs, and desires predicts morally relevant theory of mind: A longitudinal investigation. *Child Development*, *87*(4), 1221–1232. https://doi.org/10.1111/cdev.12533

Sodian, B., Taylor, C., Harris, P. L., & Perner, J. (1991). Early deception and the child's theory of mind: False trails and genuine markers. *Child Development*, *62*(3), 468. https://doi.org/10.2307/1131124

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. https://doi.org/10.1016/j.neuroimage.2007.01.042

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. https://doi.org/10.1016/j.brainres.2010.07.057

Song, H., Onishi, K. H., Baillargeon, R., & Fisher, C. (2008). Can an agent's false belief be corrected by an appropriate communication?

Psychological reasoning in 18-month-old infants. *Cognition*, *109*(3), 295–315. https://doi.org/10.1016/j.cognition.2008.08.008

Southgate, V., Chevallier, C., & Csibra, G. (2010). Seventeen-month-olds appeal to false beliefs to interpret others' referential communication. *Developmental Science*, *13*(6), 907–912. https://doi.org/10.1111/j.1467-7687.2009.00946.x

Southgate, V., Senju, A., & Csibra, G. (2007). Action anticipation through attribution of false belief by 2-year-olds. *Psychological Science*, *18*(7), 587–592. https://doi.org/10.1111/j.1467-9280.2007.01944.x

Southgate, V., & Vernetti, A. (2014). Belief-based action prediction in preverbal infants. *Cognition*, *130*(1), 1–10. https://doi.org/10.1016/j.cognition.2013.08.008

Sridharan, D., Levitin, D. J., & Menon, V. (2008). A critical role for the right fronto-insular cortex in switching between central-executive and default-mode networks. *Proceedings of the National Academy of Sciences*, *105*(34), 12569–12574. https://doi.org/10.1073/pnas.0800005105

Steffan, A., Zimmer, L., Arias-Trejo, N., Bohn, M., Dal Ben, R., Flores-Coronado, M. A., Franchin, L., Garbisch, I., Grosse Wiesmann, C., Hamlin, J. K., Havron, N., Hay, J. F., Hermansen, T. K., Jakobsen, K. V., Kalinke, S., Ko, E., Kulke, L., Mayor, J., Meristo, M., … Schuwerk, T. (2024). Validation of an open source, remote web-based eye-tracking method (WebGazer) for research in early childhood. *Infancy*, *29*(1), 31–55. https://doi.org/10.1111/infa.12564

Stets, M., & Reid, V. M. (2011). Infant ERP amplitudes change over the course of an experimental session: Implications for cognitive processes and methodology. *Brain and Development*, *33*(7), 558–568. https://doi.org/10.1016/j.braindev.2010.10.008

Stets, M., Stahl, D., & Reid, V. M. (2012). A meta-analysis investigating factors underlying attrition rates in infant ERP studies. *Developmental Neuropsychology*, *37*(3), 226–252. https://doi.org/10.1080/87565641.2012.654867

Stewart, J. L., Coan, J. A., Towers, D. N., & Allen, J. J. B. (2011). Frontal EEG asymmetry during emotional challenge differentiates individuals with and without lifetime major depressive disorder. *Journal of Affective Disorders*, *129*(1–3), 167–174. https://doi.org/10.1016/j.jad.2010.08.029

Stich, S. (1983). *From folk psychology to cognitive science*. MIT Press.

Stroganova, T. A., Nygren, G., Tsetlin, M. M., Posikera, I. N., Gillberg, C., Elam, M., & Orekhova, E. V. (2007). Abnormal EEG lateralization in boys with autism. *Clinical Neurophysiology*, *118*(8), 1842–1854. https://doi.org/10.1016/j.clinph.2007.05.005

Strommen, E. A. (1973). Verbal self-regulation in a children's game: Impulsive errors on "Simon Says." *Child Development*, *44*(4), 849. https://doi.org/10.2307/1127737

Stuss, D. T. (2001). The frontal lobes are necessary for "theory of mind." *Brain*, *124*(2), 279–286. https://doi.org/10.1093/brain/124.2.279

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

Surian, L., & Geraci, A. (2012). Where will the triangle look for it? Attributing false beliefs to a geometric shape at 17 months. *British Journal of Developmental Psychology*, *30*(1), 30–44. https://doi.org/10.1111/j.2044-835X.2011.02046.x

Surian, L., & Leslie, A. M. (1999). Competence and performance in false belief understanding: A comparison of autistic and normal 3-year-old children. *British Journal of Developmental Psychology*, *17*(1), 141–155. https://doi.org/10.1348/026151099165203

Symeonidou, M., Mizokawa, A., Kabaya, S., Doherty, M. J., & Ross, J. (2024). Contrasting one's share of the shared life space: Comparing the roles of metacognition and inhibitory control in the development of theory of mind among Scottish and Japanese children. *Developmental Science*, *27*(5), e13417. https://doi.org/10.1111/desc.13417

Taylor, M. J., & Baldeweg, T. (2002). Application of EEG, ERP and intracranial recordings to the investigation of cognitive functions in children. *Developmental Science*, *5*(3), 318–334. https://doi.org/10.1111/1467-7687.00372

Thatcher, R. W. (1992). Cyclic cortical reorganization during early childhood. *Brain and Cognition*, *20*(1), 24–50. https://doi.org/10.1016/0278-2626(92)90060-Y

Thatcher, R. W. (1994). Cyclic cortical reorganization: Origins of human cognitive development. In G. Dawson & K. W. Fischer (Eds.), *Human behavior and the developing brain* (pp. 232–266). The Guilford Press.

Thatcher, R. W., Walker, R. A., & Giudice, S. (1987). Human cerebral hemispheres develop at different rates and ages. *Science*, *236*(4805), 1110–1113. https://doi.org/10.1126/science.3576224

Thoermer, C., Sodian, B., Vuori, M., Perst, H., & Kristen, S. (2012). Continuity from an implicit to an explicit understanding of false belief from infancy to preschool age. *British Journal of Developmental Psychology*, *30*(1), 172–187. https://doi.org/10.1111/j.2044-835X.2011.02067.x

Tholen, M. G., Trautwein, F.-M., Böckler, A., Singer, T., & Kanske, P. (2020). Functional magnetic resonance imaging (fMRI) item analysis of empathy and theory of mind. *Human Brain Mapping*, *41*(10), 2611–2628. https://doi.org/10.1002/hbm.24966

Tomasello, M. (2018). How children come to understand false beliefs: A shared intentionality account. *Proceedings of the National Academy of Sciences*, *115*(34), 8491–8498. https://doi.org/10.1073/pnas.1804761115

Tullett, A. M., Harmon-Jones, E., & Inzlicht, M. (2012). Right frontal cortical asymmetry predicts empathic reactions: Support for a link between withdrawal motivation and empathy. *Psychophysiology*, *49*(8), 1145–1153. https://doi.org/10.1111/j.1469-8986.2012.01395.x

Van der Meer, L., Groenewold, N., Nolen, W., Pijnenborg, M., & Aleman, A. (2011). Inhibit yourself and understand the other: Neural basis of distinct processes underlying Theory of Mind. *NeuroImage*, *56*, 2364–2374. https://doi.org/10.1016/j.neuroimage.2011.03.053

Van Der Vinne, N., Vollebregt, M. A., Van Putten, M. J. A. M., & Arns, M. (2017). Frontal alpha asymmetry as a diagnostic marker in depression: Fact or fiction? A meta-analysis. *NeuroImage: Clinical*, *16*, 79–87. https://doi.org/10.1016/j.nicl.2017.07.006

Van Overwalle, F. (2009). Social cognition and the brain: A meta-analysis. *Human Brain Mapping*, *30*(3), 829–858. https://doi.org/10.1002/hbm.20547

Van Petten, C., & Luka, B. J. (2006). Neural localization of semantic context effects in electromagnetic and hemodynamic studies. *Brain and Language*, *97*(3), 279–293. https://doi.org/10.1016/j.bandl.2005.11.003

Vierkant, T. (2012). Self-knowledge and knowing other minds: The implicit/explicit distinction as a tool in understanding theory of mind. *British Journal of Developmental Psychology*, *30*(1), 141–155. https://doi.org/10.1111/j.2044-835X.2011.02068.x

Vogeley, K., Bussfeld, P., Newen, A., Herrmann, S., Happé, F., Falkai, P., Maier, W., Shah, N. J., Fink, G. R., & Zilles, K. (2001). Mind reading: Neural mechanisms of theory of mind and self-perspective. *NeuroImage*, *14*(1), 170–181. https://doi.org/10.1006/nimg.2001.0789

Vuga, M., Fox, N. A., F. Cohn, J., George, C. J., Levenstein, R. M., & Kovacs, M. (2006). Long-term stability of frontal electroencephalographic asymmetry in adults with a history of depression and controls. *International Journal of Psychophysiology*, *59*(2), 107–115. https://doi.org/10.1016/j.ijpsycho.2005.02.008

Vuga, M., Fox, N., Cohn, J., Kovacs, M., & George, C. (2008). Long-term stability of electroencephalographic asymmetry and power in 3 to 9 year-old children. *International Journal of Psychophysiology*, *67*(1), 70–77. https://doi.org/10.1016/j.ijpsycho.2007.10.007

Wang, L. U., & Leslie, A. M. (2016). Is implicit theory of mind the ‘real deal’? The own-belief/true-belief default in adults and young preschoolers. *Mind & Language*, *31*(2), 147–176.

Wang, Y., 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. https://doi.org/10.1016/j.neuroscience.2010.06.055

Wang, Y., Liu, Y., Gao, Y., Chen, J., Zhang, W., & Lin, C. (2008). False belief reasoning in the brain: An ERP study. *Science in China Series C: Life Sciences*, *51*(1), 72–79. https://doi.org/10.1007/s11427-008-0014-z

Wang, Y., Siu, C. T. S., & Cheung, H. (2023). P300 as a correlate of false beliefs and false statements. *Brain and Behavior*, e3021. https://doi.org/10.1002/brb3.3021

Weed, E., McGregor, W., Feldbæk Nielsen, J., Roepstorff, A., & Frith, U. (2010). Theory of mind in adults with right hemisphere damage: What's the story? *Brain and Language*, *113*(2), 65–72. https://doi.org/10.1016/j.bandl.2010.01.009

Wellman, H. M. (1990). *The child's theory of mind* (pp. xiii, 358). The MIT Press.

Wellman, H. M. (2002). Understanding the psychological world: Developing a theory of mind. In U. C. Goswami (Ed.), *Blackwell handbook of childhood cognitive development*. Blackwell Publishers.

Wellman, H. M. (2010). Developing a theory of mind. In U. Goswami (Ed.), *The Wiley-Blackwell Handbook of Childhood Cognitive Development* (pp. 258–284). Wiley Blackwell. https://doi.org/10.1002/9781444325485.ch10

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

Wellman, H. M. (2018). Theory of mind: The state of the art. *European Journal of Developmental Psychology*, *15*(6), 728–755. https://doi.org/10.1080/17405629.2018.1435413

Wellman, H. M., & Bartsch, K. (1988). Young children's reasoning about beliefs. *Cognition*, *30*(3), 239–277. https://doi.org/10.1016/0010-0277(88)90021-2

Wellman, H. M., Carey, S., Gleitman, L., Newport, E. L., & Spelke, E. S. (1990). *The child's theory of mind*. The MIT Press.

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. https://doi.org/10.1111/1467-8624.00304

Wellman, H. M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Development*, *75*(2), 523–541. https://doi.org/10.1111/j.1467-8624.2004.00691.x

Wellman, H. M., Lopez-Duran, S., LaBounty, J., & Hamilton, B. (2008). Infant attention to intentional action predicts preschool theory of mind.

Developmental Psychology, *44*(2), 618–623. https://doi.org/10.1037/0012-1649.44.2.618

Wellman, H. M., Phillips, A. T., Dunphy-Lelii, S., & LaLonde, N. (2004). Infant social attention predicts preschool social cognition. *Developmental Science*, *7*(3), 283–288. https://doi.org/10.1111/j.1467-7687.2004.00347.x

Wellman, H. M., & Woolley, J. D. (1990). From simple desires to ordinary beliefs: The early development of everyday psychology. *Cognition*, *35*(3), 245–275. https://doi.org/10.1016/0010-0277(90)90024-E

Wenzel, L., Dörrenberg, S., Proft, M., Liszkowski, U., & Rakoczy, H. (2020). Actions do not speak louder than words in an interactive false belief task. *Royal Society Open Science*, *7*(10), 191998. https://doi.org/10.1098/rsos.191998

Westley, A., Kohút, Z., & Rueschemeyer, S.-A. (2017). “I know something you don’t know”: Discourse and social context effects on the N400 in adolescents. *Journal of Experimental Child Psychology*, *164*, 45–54. https://doi.org/10.1016/j.jecp.2017.06.016

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children’s understanding of deception. *Cognition*, *13*(1), 103–128. https://doi.org/10.1016/0010-0277(83)90004-5

Winner, E., Brownell, H., Happé, F., Blum, A., & Pincus, D. (1998). Distinguishing lies from jokes: Theory of mind deficits and discourse interpretation in right hemisphere brain-damaged patients. *Brain and Language*, *62*(1), 89–106. https://doi.org/10.1006/brln.1997.1889

Woodruff, G., & Premack, D. (1979). Intentional communication in the chimpanzee: The development of deception. *Cognition*, *7*(4), 333–362. https://doi.org/10.1016/0010-0277(79)90021-0

Wysocka, J., Golec, K., Haman, M., Wolak, T., Kochański, B., & Pluta, A. (2020). Processing false beliefs in preschool children and adults: Developing a set of custom tasks to test the theory of mind in neuroimaging and behavioral research. *Frontiers in Human Neuroscience*, *14*, 119. https://doi.org/10.3389/fnhum.2020.00119

Xiao, Y., Geng, F., Riggins, T., Chen, G., & Redcay, E. (2019). Neural correlates of developing theory of mind competence in early childhood. *NeuroImage*, *184*, 707–716. https://doi.org/10.1016/j.neuroimage.2018.09.079

Xiao, Y., Zhai, H., Friederici, A. D., & Jia, F. (2016). The development of the intrinsic functional connectivity of default network subsystems from age 3 to 5. *Brain Imaging and Behavior*, *10*(1), 50–59. https://doi.org/10.1007/s11682-015-9362-z

Yamaguchi, M., Kuhlmeier, V. A., Wynn, K., & vanMarle, K. (2009). Continuity in social cognition from infancy to childhood. *Developmental Science*, *12*(5), 746–752. https://doi.org/10.1111/j.1467-7687.2008.00813.x

Yazdi, A. A., German, T. P., Defeyter, M. A., & Siegal, M. (2006). Competence and performance in belief-desire reasoning across two cultures: The truth, the whole truth and nothing but the truth about false belief? *Cognition*, *100*(2), 343–368. https://doi.org/10.1016/j.cognition.2005.05.004

Yeung, E., Müller, U., & Carpendale, J. I. M. (2019). Developmental continuity between social-cognitive skills at age 2 and false belief understanding at age 4. *Cognitive Development*, *50*, 157–166. https://doi.org/10.1016/j.cogdev.2019.05.002

Yott, J., & Poulin-Dubois, D. (2012). Breaking the rules: Do infants have a true understanding of false belief? *British Journal of Developmental Psychology*, *30*(1), 156–171.

Yott, J., & Poulin-Dubois, D. (2016). Are infants' theory-of-mind abilities well integrated? Implicit understanding of intentions, desires, and beliefs. *Journal of Cognition and Development*, *17*(5), 683–698.

Young, L., Camprodon, J. A., Hauser, M., Pascual-Leone, A., & Saxe, R. (2010). Disruption of the right temporoparietal junction with transcranial magnetic stimulation reduces the role of beliefs in moral judgments. *Proceedings of the National Academy of Sciences*, *107*(15), 6753–6758. https://doi.org/10.1073/pnas.0914826107

Young, L., Cushman, F., Hauser, M., & Saxe, R. (2007). The neural basis of the interaction between theory of mind and moral judgment.

Proceedings of the National Academy of Sciences, *104*(20), 8235–8240. https://doi.org/10.1073/pnas.0701408104

Young, L., Dodell-Feder, D., & Saxe, R. (2010). What gets the attention of the temporo-parietal junction? An fMRI investigation of attention and theory of mind. *Neuropsychologia*, *48*(9), 2658–2664. https://doi.org/10.1016/j.neuropsychologia.2010.05.012

Young, L., & Saxe, R. (2009). An fMRI Investigation of Spontaneous Mental State Inference for Moral Judgment. *Journal of Cognitive Neuroscience*, *21*(7), 1396–1405. https://doi.org/10.1162/jocn.2009.21137

Young, L., Scholz, J., & Saxe, R. (2011). Neural evidence for “intuitive prosecution”: The use of mental state information for negative moral verdicts. *Social Neuroscience*, *6*(3), 302–315. https://doi.org/10.1080/17470919.2010.529712

Yu, C.-L., & Wellman, H. M. (2024). A meta-analysis of sequences in theory-of-mind understandings: Theory of mind scale findings across different cultural contexts. *Developmental Review*, *74*, 101162. https://doi.org/10.1016/j.dr.2024.101162

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. https://doi.org/10.1016/j.neulet.2009.10.009

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. https://doi.org/10.1007/s00221-013-3661-7

Zimmer, L., Sodian, B., Mani, N., Grosso, S. S., Kristen-Antonow, S., & Schuwerk, T. (2024). *Two—To three-year-old toddlers differentiate the epistemic verbs ‘know’ and ‘think’ in a preferential looking eye-tracking paradigm*. https://doi.org/10.31234/osf.io/rby4u

Appendix

Appendix A: Translations of the example sentences

English translations of the German sentences in behavioral low-demands false belief task (Chapter 2)

German example sentence	English translation
Das ist eine Geschichte über Lily. Schau mal! Das ist Lily!	This is a story about Lily. Look! That's Lily!
Lily findet einen Apfel in einem Eimer.	Lily finds an apple in a bucket.
"Wo ist Lilys Apfel?" Links oder rechts? (an Eltern)	Where is Lily's apple?
Lily legt ihren Apfel für später in einen Korb.	Lily puts her apple in a basket for later.
Dann geht sie raus um mit einem Ball zu spielen."	Then she goes outside to play with a ball.
"Wo ist Lilys Ball?" Links oder rechts? (an Eltern)	Where is Lily's ball?
Während Lily draußen ist, findet ihr Bruder Peter den Apfel und nimmt ihn weg.	While Lily is outside, her brother Peter finds the apple and takes it away.
Lily hat Hunger. Sie kommt herein um nach ihrem Apfel suchen.	Lily is hungry. She comes in to look for her apple.
"Wo wird Lily ihren Apfel suchen?" Links oder rechts? (an Eltern)	Where will Lily look for her apple?

English translations of the German sentences in ERP belief task (i.e., Leo's belief task) (Chapter 2)

German example sentence	English translation
Schau! Das ist Leos Ball. Der Ball ist im Koffer.	Look! This is Leo's ball. The ball is in the box.
Leo denkt, der Ball ist im Koffer / Eimer.	Leo thinks the ball is in the box / bucket.
Der Eimer ist leer. Der Ball ist im Koffer.	The bucket is empty. The ball is in the box.
Wo wird Leo den Ball suchen?	Where will Leo look for the ball?

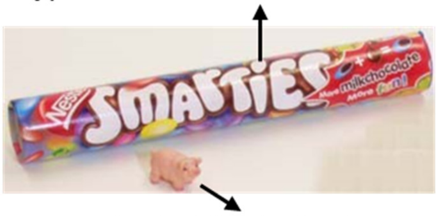
Appendix B: Procedures of the experimental tasks

Procedure of the explicit false belief task (Chapter 3)

Procedure of the content false belief task

Material: Typical Smarties box without Smarties. Toy pig that fits in the box. Figurine called Lukas.

Typical Smarties box without Smarties



Toy pig that fits in the box   Figurine called Lukas

Procedure:

Experimenter: Here is a Smarties box.

Question to the child: What do you think is in the Smarties box?

[*Get child to say "Smarties", with help if necessary. Example:*
1) *Try "Does it look like there are Smarties in it?"*
2) *Try: "What kind of box is that? What should be in it?"*
3) *Try: "Should there be Smarties in there or books?"*]

Experimenter: [*Dramatically*] Let's look… there really is a **PIG** inside the box!

[*Take out the pig and show it to the child. After a short break, put the pig back in and close the lid to prevent it from being seen.*]

Memory question to the children: Okay… what's in the Smarties box?

[If the child makes a mistake here, the content is shown again until the child gives the correct answer.]
[Figurine Lukas hidden underneath the table or in the experimenter's pocket, invisible to the child. Fetch Lukas and put him next to the Smarties box on the table so that Lukas is facing the child and introduce Lukas.]

Experimenter: This is Lukas. Lukas **still has never seen** what's in this box.

[*Then the experimenter moves Lukas towards the box so that he is now standing in front of the box and facing the box. While Lukas walks towards the box, the experimenter says:*]

Experimenter: Well, here comes Lukas.

Test question: So… what does Lukas think is in the box? Smarties or a pig? *[Repeat question if child does not answer.]*

Control question: Has Lukas ever looked in this box?

Procedure of the location false belief task

Material: Laminated sheet with a picture of a backpack and a picture of a closet. Figurine called Paul.

Figurine called Paul

Laminated sheet with a picture of a backpack and a closet

Procedure:

Experimenter: This is Paul [*place the figurine Paul next to the sheet such that Paul is approximately in the middle between the two objects*].

> Paul is looking for his gloves. Paul's gloves could be in his backpack [*point at the picture*] or in his closet [*point at the picture*].
>
> [*If the child makes their own assumption at this point, e.g. "I think the gloves are in his backpack / closet.", then proceed as follows:]*
>
> "Oh that's a good idea ... **BUT** ... **really** …" or "Oh that's a good idea …AND … really …"

Experimenter: Paul's gloves are in his backpack [*point at the picture and make a short pause*] – **but** Paul **BELIEVES** his gloves are in his closet [*point at the picture and make a short pause*].

> [*If the child does not make their own assumption and waits, just continue with the normal procedure, proceed as follows:*]

Experimenter: Well... **in reality,** Paul's gloves are in his backpack [*point at the picture and make a short pause*] - **but** Paul **BELIEVES** his gloves are in the closet [*point at the picture and make a short pause*].

Test question: Now, **where** will Paul [*point at Paul*] look for his gloves? In his backpack or in his closet?

Control question: Where are Paul's gloves **really**? Are the gloves in his backpack or in his closet?

Procedure of the day-night Stroop task (Chapter 3)

Experimenter: “Look, I brought cards with me. Look. There is a sun. When is the sun in the sky? During the day or at night? Exactly, during the day. And look, there is a moon with stars. When are the moon and stars in the sky? During the day or at night? Exactly, at night.”

<u>Introduction</u>

Experimenter: “Now we’re going to play a game with the cards! And it goes like this: “When you see a card like this [Show the child a card with the moon and stars], you have to say ‘day’. What do you have to say when you see the card?”

a) Correct answer >> “Exactly. You have to say ‘day’.”
b) Wrong answer >> “No. You have to say ‘day’.”
Experimenter: “If you see a card like this [Put the card away and show the child the card with the sun], you have to say ‘night’. What do you have to say when you see the card? “

a) Correct answer >> “Exactly. You have to say ‘night’.”
b) Wrong answer >> “No. You have to say ‘night’.”“
<u>Training</u>

Show the child a card with “a moon and stars” or “a sun”. Don’t give any instructions at first! If the child hesitates: “What do you say about the card?” (Never use the words “day” or “night” as a prompt!)

Give feedback:
a) Correct answer >> “Exactly. You have to say ‘day’/’night’.”
b) Wrong answer >> “No. On the card you have to say ‘day’/’night’.”

If there is at least one mistake after every two cards: Remind the child of the rules again using the introductory cards. Start with the card to which the child reacted incorrectly: “If you see a card like this, you have to say ‘day’; If you see a card like this, you have to say ‘night’.”

After four consecutive correct answers or after 12 cards, proceed to the test phase.

<u>16 test cards</u>

Show the child a card with “a moon and stars” or “a sun”. Don’t give any instructions at first! If the child hesitates: “What do you say about the card? “(Never use the words “day” or “night” as a prompt. No feedback!)

If the child has already answered four consecutive cards correctly during “training”, leave out the last four test cards!

Procedure of the “Simon Says” task (Chapter 3)

Instruction:

We’re going to play a game now. I’m going to show you a movement and sometimes you’ll do it too. But only sometimes: only if I say “Simon says” before the movement you join in, otherwise you don’t do anything.

For example, if I say: “Simon says: Clap your hands.” [*Experimenter claps his hands.*]

What do you do then? [*Allow time for execution or verbal description.*]

And if I just say: “Clap your hands. That is, WITHOUT saying “Simon says” first” [*Experimenter claps his hands.*]

What do you do then? [*Allow time for reaction. If action is carried out despite missing go-cue:*]

But I didn’t say “Simon says”. So again... [*Repeat until the child does it right without being prompted.*]

Before Block 2, repeat the instruction: Yes, you did a great job. And now we’ll go a little further, remember that you only ever imitate what I do when I say “Simon says”.

Block 1		**Block 2**	
Simon says	Arms up	**No-go**	Stamping
Simon says	Step forward	**Simon says**	Hands on head
No-go	Touch the nose	**No-go**	Hands on stomach
Simon says	Stamping	**Simon says**	Step back
No-go	Wave	**Simon says**	Wave
No-go	Hands on head	**No-go**	Hands on the ground
Simon says	Hands on the ground	**No-go**	Arms up
No-go	Step back	**No-go**	Step forward
Simon says	Hands on stomach	**Simon says**	Touch the nose
No-go	Grab the knee	**Simon says**	Grab the knee

Appendix C: Coding schemes

Video coding scheme for ERP belief task (Leo's belief task) (Chapter 2)

Participants
All children with ERP recording of Leo's belief task

Included segments	
Value	description
Included	Task engagement: Segments where the child maintains attention on the task presented on the monitor.
Excluded	1.Visual distraction: Children divert their gaze away from the monitor, breaking visual attention. 2.Physical movement: Segments featuring any physical movements, particularly of the head or upper body, which might disrupt EEG signal quality. 3.External interference: Parents or siblings are observed interfering with the child's engagement during the task. 4.Distraction from surroundings: Child was visibly distracted by other stimuli outside the task and does not focus on the task displayed on the monitor. 5.Lack of engagement: Child showed disinterest or an inability to continue with the task, such as persistent fidgeting or moving away from the setup.

Coding scheme of the explicit false belief task (Chapter 3)

Coding scheme of the content false belief task

Included the content false belief task	
Value	Description
0	Not included in sample if 1. Child did not reply with 'Smarties' in content belief task. 2. Child did not reply with 'Pig' in memory control question. 3. Child did not reply to one or both of the test questions. 4. Child replied ambiguously. 5. Parents or siblings interfered. 6. Experimenter made a mistake. 7. Child did not know what 'Smarties' are.
1	Included in sample.

Smarties question: Verbal utterance of the child. Should be 'Smarties' at some point! Please note if child had no content belief.	
Value	Description
No value	Child failed to reply to the question or said something else.
0	Pig(let)
1	Smarties

Memory control question: Verbal utterance of the child. Should be 'Pig(let)' at some point! Please note if child repeatedly did not pass memory control question.	
Value	Description
No value	Child failed to reply to the question or said something else.
0	Yes
1	No

Success in content false belief task	
value	description
No value	Child failed to reply to one or both of the previous questions.
0	0 in one or both of the previous questions.
1	1 in both previous questions.

Coding scheme of the location false belief task

Included the location belief task	
Value	Description
0	Not Included in sample if 1. Child did not reply to one or both of the test questions. 2. Child replied ambiguously. 3. Parents or siblings interfered. 4. Experimenter made a mistake.
1	Included in sample.

Test question	
Value	Description
No value	Child failed to reply to the question or said something else.
0	Backpack
1	Closet

Reality control question	
Value	Description
No value	Child failed to reply to the question or said something else.
0	Closet
1	Backpack

Success in content false belief task	
value	description
No value	Child failed to reply to one or both of the previous questions.
0	0 in one or both of the previous questions.
1	1 in both of the previous questions.