Continuity and Change in Theory-of-Mind Development: A Neuroscientific Approach

by

Lindsay Caroline Bowman

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Psychology)
in The University of Michigan
2013

Doctoral Committee:

Professor Henry M. Wellman, Chair
Assistant Professor Ioulia Kovelman
Associate Professor Christopher S. Monk
Assistant Professor James E. Swain

Acknowledgements

This research was funded by Rackham Graduate School, the Center for Human Growth and Development, the College of Literature, Science, and the Arts, and the Department of Psychology; all of the University of Michigan. This research was also funded by the Eunice Kennedy Shriver National Institute for Child Health and Human Development, and the Natural Sciences and Engineering Council of Canada.

I would like to thank the hardworking parents and children who participated in these studies, and the dedicated research assistants who helped with data collection and analysis. I would also like to thank my graduate advisor for his invaluable advice and contributions, as well as all members of my dissertation committee for their time, guidance, and helpful feedback. Finally, I thank my friends and family for their endless support.

Table of Contents

List of Tables

Table

List of Figures

Figure

Abstract

This dissertation seeks a more comprehensive picture of the process of theory-of-mind development (an understanding that action is guided by internal mental states such as beliefs, desires, and intentions) by examining *neurological correlates* of theory-of-mind reasoning in child populations. Taken together, three studies, each using a different neuroscientific method, span investigations of early, middle, and late childhood, substantially increasing the currently small pool of pediatric neurocognitive research. The studies highlight the utility of a neuroscientific approach to theory-of-mind investigations, and illustrate the importance of collecting *developmental* data. Results exemplify how, when applied in the developmental context, neuroscientific methods shed light on open questions of continuity and change in theory-of-mind development, and how different mental-state understandings (e.g., beliefs and desires) build to form a complete theory of mind.

Study 1 uses fMRI and resting EEG methods to track theory-of-mind neural correlates longitudinally from 4- to 7-years-old. Results provide direct evidence for continuity and indirect evidence for change in the relation between theory-of-mind reasoning and neural substrates over early to middle childhood. Moreover, they provide validation for novel methodology that can advance theory-of-mind research in younger, and social-cognitively impaired populations. Study 2 uses fNIRS to investigate belief- and desire-reasoning in 7- and 8-year-olds. Results suggest a mechanism by which belief-understanding may build off prior desire-understanding, and demonstrate a method by which changes in these mechanisms can be measured as a function of performance accuracy. Study 3 uses ERP to examine neural correlates of belief- and desire-reasoning in 10- and 11-year-olds. Results provide important developmental comparisons to two prior, parallel ERP studies, to further illuminate the complex neurodevelopmental processes underlying advancement in belief- and desire-reasoning from middle to late childhood.

These studies allow cross-methodological comparisons, combining strengths to overcome weaknesses of any single method, and provide test cases for adapting neuroscientific methods for children. Theoretically, they pioneer in forming a comprehensive picture of how theory-of-mind mechanisms and conceptions change with time and experience, and further illuminate the processes governing theory of mind and its development.

CHAPTER I

General Introduction

Humans learn to navigate a complex social world. We communicate, cooperate, empathize with, and make sense of others' behaviors. In short, we develop *social cognition.* A cornerstone of social cognition is the development of *theory of mind*—the understanding that human behavior is guided by internal mental states (Harris, 2006; Wellman, 2011). Importantly, theory of mind goes beyond social perception; it requires conceptualization of and reasoning about people's mental states in order to accurately predict and explain behavior. Though sometimes equated with children's achievement of understanding false beliefs, I advocate a much broader construal of theory of mind, both conceptually and developmentally. Theory of mind describes the wide-ranging human understanding of agents' intentions, desires, and thoughts, and how each of these internal states ultimately shape action. It appropriately emphasizes the mentalism that so strongly characterizes our everyday psychological understanding of the social world.

Understanding how theory of mind develops has importance for uncovering fundamentals of everyday social cognition, and for understanding development in populations marked by social-cognitive impairment as in individuals with autism and deaf children born to non-signing parents (e.g., Baron-Cohen 2001; Peterson, 2004; Peterson, Wellman & Liu, 2005). Researchers have investigated theory-of-mind behavioral developments in infants (e.g., Onishi & Baillargeon, 2005), young children (e.g., Wellman, Cross & Watson, 2001), older children (e.g., Friedman & Leslie, 2004) and adults (e.g., Keysar, Lin & Barr, 2003). These investigations include both typically developing and atypically developing (e.g., autistic, deaf) individuals (e.g., Peterson et al., 2005) across various cultures (e.g., Callaghan et al., 2005). Indeed, behavioral research frames our current understanding of changes and advances in theory-of-mind reasoning. Yet, there are limits to what behavioral research alone can teach us. Currently, the mechanisms and processes underlying theory-of-mind development are still unclear.

Cognitive neuroscientific methods can shed light on the mechanisms and processes that underlie, influence, and pace development. Identification of the neurological correlates of cognition (neural substrates, networks, and signatures that are functionally related to cognitive processing) provide insight into cognitive development generally, and theory of mind specifically, when behavioral measures are limited or ambiguous. A wealth of neuroscientific research has now identified specific neural substrates consistently recruited for belief and false-belief reasoning in adults (see e.g., Carrington & Bailey 2009, Apperly, 2011). This adult neurocognitive research frames initial understanding of theory-of-mind neural correlates, but it also leaves much unknown about *development*. Adult data alone cannot capture the developments that lead to mature achievements, and from a behavioral perspective theory of mind develops dramatically (e.g., Wellman et al., 2001; Sodian 2011). Currently, there are very few investigations of theory-of-mind neural correlates in children.

The three studies in this dissertation use cognitive neuroscientific methods to investigate theory-of-mind development in child populations. Thus, these studies extend the adult findings to examine *developmental* issues by directly investigating neural substrates associated with theory of mind as it develops from the preschool years into middle and late childhood, greatly enhancing the currently small pool of existing pediatric neuroscientific investigations of theory of mind. Moreover, they address central questions left unanswered by the behavioral data alone. Thus, taken together, these three studies bridge gaps in both behavioral and neurocognitive theory-of-mind research, and work to capture a more complete picture of the processes of theory-of-mind development.

In the following sections of this chapter, I summarize key behavioral research that lays foundation for understanding theory-of-mind reasoning, but also highlights key unanswered questions central to theory-of-mind development. I briefly outline how cognitive neuroscientific research on theory of mind in adults has begun to address some of these questions, yet also demonstrates a need to address development further, through direct examinations of neural correlates of theory of mind in children. I then highlight some of the existing pediatric neurocognitive investigations of theory of mind, which demonstrate the importance of child investigations, yet only begin to address outstanding questions of development. I conclude this chapter with a brief preview of the three studies that comprise this dissertation, highlighting their key methodological and theoretical contributions to the study of theory-of-mind development.

Behavioral Research: Framing Developmental Understanding and Fostering New Neuroscientific Questions

Behaviorally, theory of mind is a deeply developmental achievement. Outlining these behavioral developments frames current understanding, but also highlights unanswered questions and new research for which cognitive neuroscientific methods offer a valuable approach.

Infancy

Over the first year of life, infants show increased complexity in their understanding of others and others' mental states. By 3 months of age, infants follow another's eye gaze (e.g., D'Entremont, Hains, & Muir, 1997) and engage in joint visual attention in social contexts (Tremblay & Rovira, 2007). By 12-14 months, infants also follow an adult's gaze around a barrier, even if this requires leaning or moving behind the barrier on their own, demonstrating initial insights into the notion that experience is subjective (e.g., someone might not see what I see/someone might see something I do not) (e.g., Moll & Tomasello, 2004). By 8 to 9 months, infants can identify the goals and intentions behind others' successful actions—i.e., actions that culminate in an obvious achieved goal such as reaching for and retrieving a ball (Woodward 1998). But by 10-12 months, infants have a more solid understanding of intentionality, and can identify the intentions behind 'failed' actions as well—i.e., actions that do not culminate in achieved goals, such as reaching towards but never grasping a ball behind a barrier (Brandone & Wellman, 2009).

Infants develop in their understanding of others' desires as well. At 12 months of age, infants understand that actions are based on desires (e.g., they show surprise when an actor chooses a toy that does not match her expressed desire) (Phillips, Wellman, & Spelke, 2002). But beyond 12 months, infants further recognize that desires are subjective and distinct; that they are specific to individuals and can sometimes contrast with infants' own desires. Thus, 18-month-olds will offer an experimenter a snack for which she previously expressed preference (i.e., "Mmm, Yummy!" after eating broccoli) even if the experimenter's desire contrasts with the infants' preference for a different snack (i.e., crackers) (Repacholi & Gopnik, 1997).

Some research also suggests that older infants possess an implicit (non-verbal) understanding of beliefs. Investigations of implicit understanding typically measure infants' eye gaze (e.g., length of gaze, anticipatory saccades) as a way of uncovering infants' underlying expectations about the scenes they are viewing. In an initial, often-cited study, Onishi & Baillargeon (2005) found that 15-month-olds look longer when an actor behaves in a way that is inconsistent with her prior belief (e.g., searches for a toy in location B when she originally placed the object in location A and did not see the object move). The authors argue that this longer looking-time demonstrates that infants have an expectation that actions are guided by beliefs, and thus they have an implicit form of false-belief understanding. This finding has been replicated with other looking-time studies (Surian, Caldi & Sperber, 2007) as well as in research using eye-tracking measurements (e.g., Southgate, Senju & Csibra, 2007).

The above collection of findings taken together demonstrates development of mental-state understanding over the first two years of life: infants' understanding advances from initial unsteady links between action and intention, to more solid conceptions of person-specific desires and, perhaps, an implicit understanding of beliefs. However, an open question concerns how these implicit, non-verbal forms of understanding relate to the explicit, verbal understanding demonstrated in older children and adults. Indeed, there is much debate surrounding the interpretation of the infant false-belief studies. Some scholars argue implicit false-belief tasks could be solved with rule-based strategies that do not rely on belief-understanding (e.g., Perner & Ruffman, 2005), or with simple registration of others' engagement with objects, which is merely "belief-like" (Apperly & Butterfill, 2009). Thus, though much is known about theory-of-mind development in infancy, many questions remain. In particular, there is much debate over whether implicit false-belief tasks capture true development of belief-understanding in infancy.

Later Developments

In contrast, extensive investigation of behavioral theory-of-mind developments beyond infancy demonstrates clear and robust advances in belief-understanding. Indeed, hundreds of studies have been conducted using variations of the 'false-belief task' (see Figure 1.1) across many different cultures (e.g., Callaghan et al., 2005). These studies all demonstrate that children transition from consistently failing explicit false-belief tasks at around age 3 years, to consistently passing them by around age 5 years (see Wellman et al., 2001 for meta-analysis).

So much attention has been paid to this robust achievement, that the passing of the false-belief task is sometimes equated with 'achievement of theory of mind'. However, focus on a single task-type obscures crucial developmental aspects of theory of mind; theory of mind involves understanding multiple interconnected mental concepts and developmentally, children's theory of mind proceeds in a progression of mental-state understandings. Indeed, a crucial, well-documented progression is that children consistently develop an explicit understanding of desires before developing an explicit understanding of beliefs. Although children use desire words (e.g., 'want', 'like') and can predict that storybook characters will act according to their desires by around age 2 years, it is not until around 3 years of age that they use belief words (e.g., 'think', 'know') and consistently make predictions for characters based on beliefs (Bartsch & Wellman, 1995; Wellman & Woolley, 1990). Moreover, the desire-belief progression holds across tasks matched on procedural methodology, linguistic structure, and materials (e.g., Wellman & Liu, 2004).

As an example of one well-controlled contrast, consider diverse-desires versus diverse-beliefs tasks. In these tasks, children are told about a character that likes (for the diverse-desires tasks) or thinks (for the diverse-beliefs tasks) something opposite to what the child thinks (e.g., child likes cookies but character likes carrots/child thinks cat is in tree but character thinks cat is in garage). Children are then asked to predict the character's actions (e.g., what will the character choose to eat?/where will the character look for his cat?). Though the demands and format for the two tasks are virtually identical, children consistently pass diverse-desires tasks at an earlier age than diverse-beliefs tasks (e.g., Wellman & Liu, 2004; Wellman & Woolley, 1990). Indeed, the desire-belief progression is present across different cultures (e.g., Wellman, Fang, Liu, Zhu & Liu, 2006), and it even occurs in populations in which social development is impaired (e.g., in late-signing deaf children and high functioning individuals with autism; Peterson et al., 2005).

However, despite the wealth of converging evidence demonstrating advances in mental-state understanding from early to middle childhood, the nature of this cognitive-developmental progression remains unclear from behavioral research alone. For example, children's performance reaches ceiling on the standard behavioral theory-of-mind tasks at roughly age 6 or 7 years (see e.g., Wellman et al., 2001; Wellman & Liu, 2004), yet belief- and desire-reasoning surely develop beyond these ages. Moreover, several possible underlying mechanisms can predict or account for the same sequences of development. Indeed, there exist today several alternative accounts of theory of mind and its development, and most of them fall into two competing camps. One camp emphasizes the role of innate “theory-of-mind modules” that are in place in infancy and that gradually come on-line over the course of development. These accounts posit that development of mental-state concepts is relatively uniform and stable over time and across populations (e.g., Scholl & Leslie, 1999). Theory theories, in contrast, posit that children develop internal naïve theories about the mind and about mental states, and that theory-of-mind understanding builds gradually as these internal naïve theories are revised and changed based on new experience (e.g., Gopnik & Wellman, 1992; 2012).

Open Questions for Neuroscientific Research

Fundamental questions of both developmental *continuity* and *change* thus remain unsettled in theory-of-mind research. The application of neuroscientific methods provides a promising addition to behavioral investigations in service of investigating these questions. As demonstrated in adults, identification of neural correlates can highlight component processes (both domain-general and -specific) that underlie complex cognitive reasoning (e.g., Saxe, Schulz, & Jiang, 2006 and Saxe & Powell, 2006). And they can reveal underlying similarities where behavioral data show differences, as well as underlying differences in processing where behavioral data show performance similarities (e.g., Liu, Meltzoff, & Wellman, 2009a).

When neuroscientific methods are used in the developmental context, they can be even more valuable. Indeed, though cognitive neuroscience has been criticized both as reductionistic (by many psychologists), and as simplistic (by cellular neuroscientists), its intermediate position can provide a level of analysis that helps clarify developmental continuity and change. For example, identifying the neural correlates of theory of mind in infants and young children can reveal processes underlying cognition at the outset of development, when behavioral methods and data are limited. These early correlates can be compared with neural mechanisms present in expert adults to investigate differences and similarities pointing to developmental continuity and change. In contrast, behavioral methods spanning this wide age range are so discrepant that they obscure developmental comparisons. Neural mechanisms underlying implicit reasoning can be compared with those for explicit reasoning to address questions of how infant mental-state understanding relates to the understanding demonstrated in older children and adults. Further, tracking these neural correlates across development can potentially identify internal developments even prior to behaviorally manifested change. And as neural correlates for typical theory-of-mind development are identified, they can be compared to populations exhibiting theory-of-mind delays. Thus, cognitive neuroscientific investigations of theory of mind that directly examine child populations can shed light on how cognition along with its correlated underlying neural processes grow and change with experience, between populations, and across the lifespan.

Neural Correlates of Theory of Mind in Adults

Currently, there are few neuroscientific investigations that examine theory of mind in young children directly. However, abundant neurocognitive studies on theory of mind in adults lay important foundation for developmental research. Converging evidence demonstrates that for adults, theory-of-mind reasoning consistently recruits a network of specific neural regions (Figure 1.2): the medial prefrontal cortex (MPFC) and the left and right temporoparietal junction (TPJ) most consistently, as well as the superior temporal gyrus/sulcus (STG/S), the precuneus, and the temporal poles (see Apperly, 2011; Carrington & Bailey, 2009; for recent reviews review). These regions are recruited when adults engage in multiple mental- and social-reasoning tasks across functional neuroimaging (fMRI) and electrophysiological (EEG/ERP) methods alike.

To illustrate, consider adult decoding of mental states. Adults showed increased BOLD (blood-oxygen-level-dependent) signal (the hemodynamic response that indexes neural activation and is tracked by fMRI) in the left PFC, the medial frontal gyrus, and the left STG when inferring mental states from photographs of eyes (e.g., desirous, thoughtful, confused) versus determining the gender of the eyes (e.g., Baron-Cohen et al., 1999; Adams et al., 2009). When this task was adapted to ERP methods, adults showed increased electrophysiological activity in prefrontal and medial temporal cortex (e.g., Sabbagh, Moulson, & Harkness, 2004).

Beyond mental-state decoding of simple static images, theory-of-mind neural regions are recruited when processing descriptions of more complex social interactions and scenes. Adults showed increased BOLD signal in the medial frontal cortex and the anterior cingulate cortex when reading stories requiring general mental-state inferences (i.e., inferring thoughts and knowledge states from a series of human actions and interactions) compared to reading stories requiring non-mental mechanical inferences (i.e., deducing why an alarm sounded) (Vogeley et al., 2001). Contrasts between mental-state descriptions and non-mental *human* descriptions (e.g., non-mental social interactions and descriptions of peoples' appearances) revealed activity in adults' bilateral TPJ, anterior STS, and medial frontal gyrus (Saxe & Kanwisher 2003).

Many adult neurocognitive studies focus on identifying neural correlates of belief- and false-belief reasoning specifically. This is likely because, from a behavioral perspective, achievement of explicit false-belief understanding is often marked as the 'gold standard' for developing a theory of mind. Considerable evidence demonstrates that belief-reasoning activates the theory-of-mind neural network. For example, when attributing true and false beliefs to cartoon characters, adults showed activation in the superior, middle, and inferior frontal gyrus, the left and right inferior parietal lobule, and the precuneus. The TPJ, prefrontal cortex, and precuneus showed more activation to false-belief attribution compared to attribution of true beliefs (Sommer et al., 2007). Similarly, in an ERP study, Liu et al. (2009a) found that when adults reasoned specifically about beliefs, they showed electrophysiological activation in both mid-frontal and right-posterior scalp regions.

There is some converging evidence that belief-reasoning may be specialized to posterior brain regions (i.e., TPJ). Saxe and Wexler (2005) demonstrated that adults selectively recruited the right TPJ for processing mental states but not for processing other socially relevant facts about a person (i.e., marital status, family relations, cultural background). Moreover, none of the other regions typically recruited in theory-of-mind reasoning (i.e., medial prefrontal cortex) showed such a specified role. Further, Saxe and Powell (2006) found that activations in the TPJ and posterior cingulate were selectively associated with reasoning about beliefs but not with reasoning about other socially relevant facts such as a person's appearance, or about other specific non-mental internal states such as bodily sensations. Some evidence suggests that the *right* TPJ in particular is specific for belief-reasoning, above and beyond any involvement in domain-general computations that are also important for theory-of-mind such as executive functioning skills (i.e., inhibitory control, attention attribution, working memory). In their fMRI study, Saxe et al. (2006) found that brain regions recruited during an executive functioning task (bilateral intraparietal sulcus, frontal operculum, middle frontal gyrus, and middle temporal gyrus) did not overlap with any regions recruited for a belief-attribution task (left and right TPJ, MPFC, and anterior STS). A third task that required both executive functioning *and* belief-attribution skills did recruit several of the same neural regions as recruited in the executive functioning task; however, the combined task also selectively recruited the right TPJ, which was *not* recruited in the executive-functioning task. According to the authors, this pattern of results suggests there are distinct, domain-specific cognitive processes for belief-reasoning, and that the right TPJ may be specialized for this domain-specific reasoning.

Converging evidence consistently demonstrating a network of regions, and potentially demonstrating further specialization of particular regions within the network, has important implications for developmental research. Identification of these kinds of 'neural signatures' can be useful for addressing larger and more complex questions of cognitive processing and cognitive development, and it can also shed indirect light on developmental issues. As one example, findings from the Saxe et al. (2006) study investigating belief-reasoning and executive functioning described above sheds some light on the role of executive functioning in theory-of-mind development. That is, one account of theory-of-mind development posits that advances in mental-state understanding represent domain-general maturation of executive functioning skills; these domain-general skills help children navigate the computational and task demands of mental-state attribution (e.g., Carlson & Moses, 2001). However, the dissociable substrates for belief-reasoning and executive functioning reported in Saxe et al. provide evidence that not all theory-of-mind reasoning is governed by executive functioning—some substrates show a domain-specific role for belief-reasoning beyond any involvement in executive functioning.

Limits of Adult Neuroscientific Investigations

However, there are limits to what the current adult neurocognitive research can teach us about theory-of-mind development. As one example, theory of mind encompasses understanding of and reasoning about multiple distinct mental states, including intentions, desires and emotions; yet, current adult neuroscientific studies typically examine mental-state reasoning generally, or even more commonly, belief-reasoning by itself.

Consider an fMRI study by Saxe and Kanwisher (2003) that primarily examined false-belief reasoning by contrasting it with reasoning about non-mental representations (i.e., false photographs). Indeed, TPJ regions were involved in false-belief reasoning in contrast to false-photograph reasoning. In this study Saxe and Kanwisher also included mental stories about desires and briefly reported that TPJ regions involved in false-belief reasoning also responded significantly to desire-reasoning. But the data for processing desires specifically were not mentioned or analyzed further. Thus, there are almost no data available on the neural correlates of reasoning about desires versus beliefs, and none comparing responses on strictly matched mental-state tasks.

Neuroscientific investigations that directly contrast desire-reasoning and belief-reasoning seem particularly important from a developmental perspective. Behavioral research clearly demonstrates that development of desire-understanding precedes development of understanding beliefs (e.g., Wellman & Liu, 2004; Wellman et al., 2006). Yet the mechanisms underlying this developmental progression are unknown. Neuroscientific methods could shed light on this issue, and further identification of neural correlates of desire-reasoning compared to belief-reasoning could provide a window on the neural mechanisms that more generally support understanding of *different* mental states.

A recent ERP study with adults demonstrates the utility of including direct belief-desire contrasts. Liu et al. (2009a) recorded ERPs as adults performed diverse-desires tasks (requiring reasoning that different persons can have different desires for exactly the same thing) and diverse-beliefs (requiring reasoning that different persons can have different beliefs about the exact same situation). As a control, participants performed parallel diverse-physical tasks (requiring reasoning about where different things go). A late slow wave (LSW) with mid-frontal scalp distribution was associated with desire- *and* belief-judgments. However, an LSW with right-posterior scalp distribution was associated *only* with belief-judgments. This neural dissociation is even more striking considering behavioral performance accuracy across the three conditions was identical. Note, at a broad level, these results demonstrate how neurocognitive data can provide distinctive information when behavioral data do not; they revealed underlying neural differences despite behavioral equivalences. More specifically, these findings demonstrate neural overlap as well as critical differences in reasoning about desires and beliefs, and point to a possible explanation for the progression of understanding desires to understanding beliefs seen in children. Children may need to recruit additional neural processes (within posterior parietal regions) for reasoning about beliefs, beyond a common neural system (within medial frontal regions) for reasoning about mental states more generally.

Only one other neuroscientific study with adults that I know of has directly compared desire- versus belief-reasoning (Abraham et al., 2010). Thus, it is difficult to definitively conclude that right posterior parietal regions (e.g., right TPJ) are specifically specialized for belief-reasoning, over and above recruitment of other mental states such as desires. Indeed, the role of the right TPJ as specific for beliefs, or even as specific for mental-state reasoning more generally, is currently debated. In contrast to Saxe and colleagues, several researchers argue that the right TPJ is recruited for theory-of-mind reasoning due to a domain-general role in attention shifting, rather than a domain-specific role in processing beliefs (see Mitchell, 2008; Rothmayr et al., 2011). Thus, the current adult literature is far from conclusive; it raises as many questions as it answers about the neural mechanisms underlying development. For example, it leaves specific questions about mechanisms underlying the developmental progression from desire-understanding to understanding beliefs.

Of course, even if the adult findings were clearer, it is not possible to fully address *developmental* issues from adult data alone. An understanding of neurocognitive correlates in cognitively expert adults does not translate to an understanding of cognition earlier in development. Clearly, there are many reasons to think that adult and child cognition differs considerably due to differences in experience, domain general capabilities, and cognitive strategies (Karmiloff-Smith, 1997). The methods and findings from the adult literature provide an important foundation to launch pediatric examinations of theory-of-mind neural correlates. And it is these direct neurocognitive examinations of child populations that are needed to more fully address important and outstanding questions in theory-of-mind development.

Neural Correlates of Theory of Mind in Children

Recently, researchers have begun to examine the neural correlates of theory of mind in children. A more extensive review of existing pediatric neurocognitive investigations of theory of mind is presented in the introduction of Study 1, directly following this chapter. Below, I highlight only a few examples that demonstrate the importance of these kinds of pediatric investigations for uncovering processes underlying theory-of-mind development.

In an early investigation of children's theory-of-mind neural correlates, Ohnishi and colleagues (2004) used fMRI to examine action-intention understanding and mental-state attribution in typically-developing children aged 7- to 13-years-old. By comparing neural activation in action-intention conditions (viewing purposeful hand actions such as grasping a cup) to neural activation in mental-state attribution conditions (viewing triangles moving in 'mentalistic' ways such as one triangle 'jumping out to surprise' another triangle), the authors investigated the relation between intention-understanding and 'higher level' theory of mind. Results showed that the action-intention and mental-state attribution conditions both yielded neural activation in the bilateral STS, temporal lobes, and fusiform gyrus. However, activation in the right MPFC, the right inferior parietal cortex, and the right TPJ was unique to the mental-state attribution condition.

According to the authors, this pattern of overlap and distinction could suggest that theory of mind develops from an initial capacity to detect biological motion, and subsequently to infer intentions from actions (processing which would occur in the STS and fusiform gyrus). Additional substrates would then be recruited (in the MPFC and right TPJ) to support perspective-taking, facilitating more complex mental-state reasoning.

Mosconi and colleagues (2005) similarly used fMRI to examine intention-understanding in typically developing children 7- to 10-years old. Intention-understanding measures for this study consisted of viewing a character shift her eye-gaze toward a target image (an action with a clear intention of viewing a target object), contrasted with viewing the same character shift her eyes away from the target image toward an empty space (an action with a less-clear goal). The STS, middle temporal lobe, and inferior parietal lobule showed increased activation in the 'shift-toward' condition compared to the 'shift-away' condition. Thus, this study provides converging evidence that the STS is recruited for intention-understanding (this time measured via eye-gaze) in children as young as 7 years of age, and provides additional support for the developmental hypothesis from Ohnishi et al. (2004) outlined above.

There is also some research that measures neural correlates in even younger children, at points in development when behavioral tasks demonstrate stark, outward change (i.e., transitioning from failing to passing standard false-belief tasks). Sabbagh, Bowman, Evraire, and Ito (2009) used dense-array (128-channel) EEG recordings to investigate how 4-year-olds' resting EEG alpha coherence (a measure of functional maturation of underlying neurocognitive systems; e.g., Nunez, 1995) related to their theory-of-mind development (e.g., performance on appearance-reality and false-belief tasks). Children's executive functioning performance (e.g., response-conflict, inhibition) was measured as a covariate. Source localization of the EEG alpha showed that increased functional maturation of the dorsal MPFC and the right TPJ predicted increased theory-of-mind performance in these children. Importantly, these relations held even after statistically controlling for children's executive functioning performance. Thus, these child data shed even more light on the question of whether advances in mental-state understanding can be largely accounted for by domain-general maturation of executive functioning skills (e.g., Carlson & Moses, 2001). Specifically, the child data from Sabbagh et al. extend similar findings in adults (Saxe et al., 2006) to provide evidence that even early in development, theory of mind includes both domain-general and domain-*specific* substrates (e.g., substrates that correlate with theory-of-mind development independent of any common relation with executive functioning), demonstrating that developments in executive functioning cannot entirely account for development of theory of mind, even at these younger ages when children exhibit much outward behavioral advancement. Moreover, these child data demonstrate that domain-specific substrates exist not only in the right TPJ (as in the adult research; Saxe et al., 2006), but in the MPFC as well (a region that did not show domain-specific processing in adults), suggesting that neurocognitive systems supporting theory-of-mind reasoning in adults may not be entirely the same as the systems that support children's initial theory-of-mind reasoning, but rather may be the product of some developmental reorganization and change.

Just these few examples illustrate the importance of pediatric investigations of theory-of-mind neural correlates for extending adult neurocognitive research. Such an extension more directly addresses developmental hypotheses by examining the neurological processes underlying theory of mind as it develops in children. Currently though, the pool of these pediatric investigations is small, and together existing studies provide only brief snapshots of neurocognitive correlates at select ages, creating a picture of development that is still patchy and full of unknowns. Longitudinal data that track continuity and change in theory-of-mind neural correlates across age are needed to more directly clarify the developmental picture, yet currently no such data exist. Moreover, just as in the neuroscientific studies of adults, the existing child studies have narrowly focused on belief-reasoning, or on mental-state reasoning in general. To reiterate, because behavioral data clearly demonstrate distinct developments for different types of mental states (e.g., intentions, desires, beliefs), focus on either undifferentiated mental states generically, or on beliefs and false-beliefs alone is limited. The most prominent and substantiated developmental progression is one from robust early desire-understanding to later belief-understanding (e.g., Wellman & Liu, 2004); thus child neuroscientific investigations should address multiple mental-state understandings, and critically, *desires* as well as beliefs.

As I outline next, the three studies of the dissertation address these most outstanding issues, providing both needed longitudinal data, and data contrasting neural correlates for belief- and desire-reasoning, in child populations. Taken together, they demonstrate useful methods, offer valuable clarity on several outstanding developmental questions, and point to further, feasible research.

The Dissertation

This dissertation employs a suite of neuroscientific methods to investigate the mechanisms underlying theory-of-mind development through direct examination of theory-of-mind neural correlates over early, middle, and late childhood. The dissertation is in multiple manuscripts form, comprising three studies that both extend behavioral and adult neuroscientific research, as well as greatly enlarge the currently small pool of existing research on the neural correlates of theory of mind in children. Moreover, taken together, these studies address three outstanding questions central to an understanding of theory-of-mind development: 1) How does theory of mind continue to develop beyond early childhood? 2) To what extent is later development different for differing mental-state concepts (i.e., for desires versus beliefs)? 3) What factors contribute to children's progression from understanding desires to understanding beliefs?

Study 1 uses functional magnetic resonance imaging (fMRI) to track neural correlates of theory of mind longitudinally from 4- to 7-years-old, and investigates whether neural specializations for theory-of-mind reasoning change as children grow. This study is a direct follow-up to a previous study that investigated neural correlates of theory of mind in preschoolers using source-localized electroencephalography (EEG) (Sabbagh et al., 2009). Study 2 uses functional near-infrared spectroscopy (fNIRS) to investigate whether 7- and 8-year-olds recruit neural substrates differently for reasoning about beliefs versus desires, and suggests a mechanism by which belief-understanding may build off prior desire-understanding. This study is a direct follow-up to a recent study that used ERP methods to examine neural correlates of belief- and desire-reasoning in 7- and 8-year-olds (my 619 research; Bowman, Liu, Meltzoff, & Wellman, 2012).

Finally, Study 3 uses event-related electrophysiology (ERP) to examine neural correlates of belief- and desire-reasoning in 10- and 11-year-olds to further illuminate the process by which mental-state understandings build to form a complete and expert theory of mind.

These studies make both methodological and theoretical contributions. In employing multiple neuroscientific methods (fMRI, fNIRS, and ERP), the studies allow important cross-methodological comparisons, combining strengths to overcome inevitable weaknesses of any single method. They also provide initial validation and standardization of novel neuroscientific paradigms. Moreover, they provide valuable methods for adapting existing neuroscientific approaches for use with young children. Theoretically, these studies advance understanding of two central concepts of theory-of-mind development. Most specifically, these studies shed light on the processes underlying the progression from desire- to belief-understanding. As noted earlier, both behavioral and neuroscientific investigations have focused primarily on belief- and false-belief understanding in children and adults. However, focus on beliefs and false beliefs alone is limited; Theory of mind—often termed a belief-desire or belief-desire-emotion naïve psychology—involves understanding multiple, causally interconnected mental concepts, as evidenced most strongly by the robust progression from understanding desires to understanding beliefs (e.g., Wellman & Liu, 2004). Studies 2 and 3 include a comparison of the neural correlates of belief- *and* desire-reasoning in the same sample of children, directly measuring similarities and differences in underlying developmental processes supporting reasoning about these two mental states. Thus, these studies provide a more comprehensive examination of how different mental-state concepts develop to form a complete theory of mind.

More broadly, the studies in this dissertation shed light on continuity and change in theory-of-mind development—one of the most fundamental and yet under-explored issues in the theory-of-mind development literature. Study 1 tracks the neural correlates of theory of mind longitudinally, and provides a direct measure of continuity and change of underlying neural processes as children grow from early to middle childhood. Moreover, across all three studies, identification of the *neurological correlates* of theory-of-mind reasoning in childhood allows a more fine-grained window on continuity and change in developmental processes compared to behavioral data alone. As mentioned previously, identification of neural processes can reveal similarities over time/across populations where behavioral data show differences, as well as underlying differences in processing where behavioral data show performance similarities (e.g., Liu et al., 2009a). Moreover, these early correlates can be compared with neural mechanisms present in expert adults to investigate differences and similarities pointing to developmental continuity and change. Tracking these neural correlates across development can identify internal developments prior to behaviorally manifested change, and give further clarity where behavioral measures are limited or ambiguous. Thus, the three studies, taken together, help to build a more complete conceptualization of continuity and change—capturing a more comprehensive picture of the process of theory-of-mind development itself.

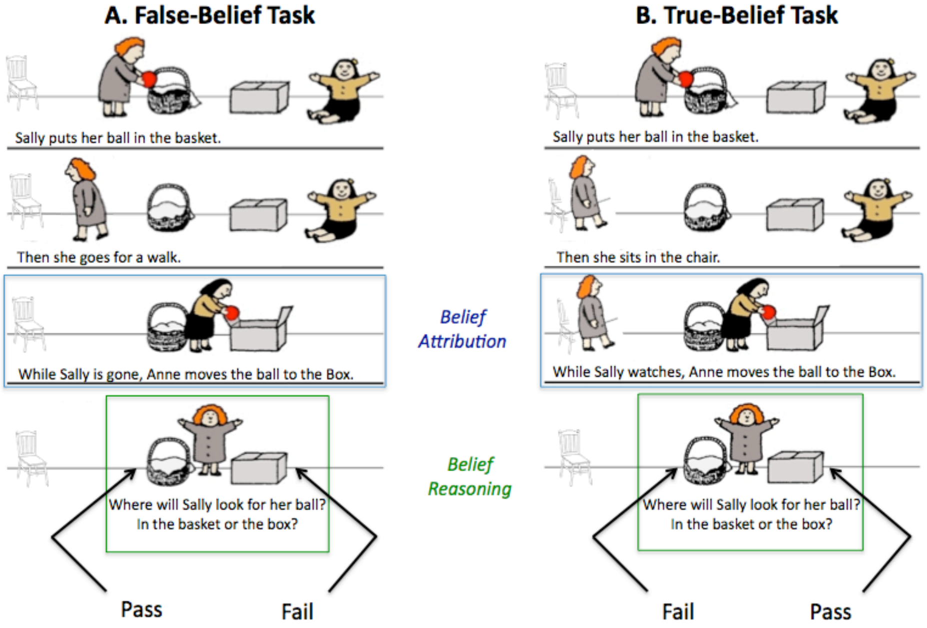


Figure 1.1. Prototypical examples of false-belief (A) and true-belief (B) tasks. In each case, participants are required to attribute a belief (a false belief in A, but true belief in B) to Sally regarding the location of her ball, and then to reason about Sally's actions based on that attribution. To pass the false-belief task, participants must answer that Sally will look for her ball in the basket. This reasoning requires understanding that mental-states guide action (Sally should think the ball is still in the basket, and thus search for it there), and moreover that these mental states are both person-specific (even though Anne and the participant know the location of the ball, Sally does not) and distinct from reality (even though the ball is truly in the box, Sally's search is driven by her belief that it is in the basket). Participants fail the task when they answer that Sally will look for her ball in the box, thereby demonstrating an incomplete understanding of beliefs that is confounded by reality, and/or by others' mental states. Extensive behavioral research demonstrates that typically developing children consistently fail this task around ages 2 and 3 years, but pass it by around age 5 years. True-belief tasks require similar reasoning, though they do not fully separate mental-states from reality and other people, and thus passing the false-belief task is considered the gold standard for achieving full belief-understanding. Given the complexities of the mental concepts required in the false-belief task, passing of this task is often equated with achievement of theory of mind more generally.

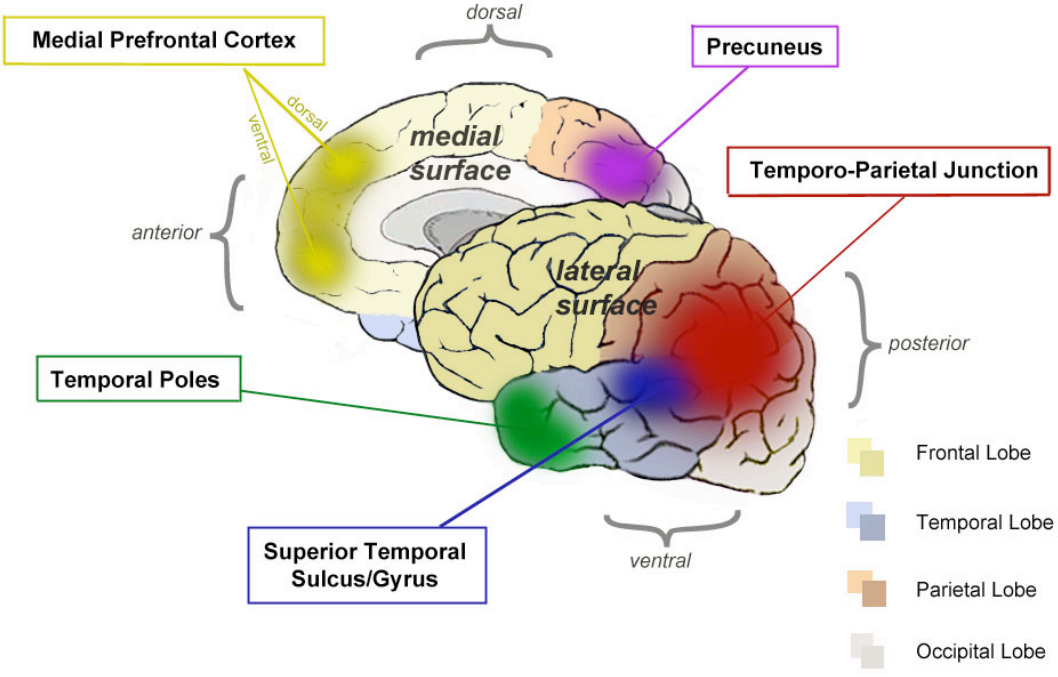


Figure 1.2. Depiction of the neural regions comprising the theory-of-mind network. Though shown on only one hemisphere, research demonstrates that each of these regions is recruited bilaterally for theory-of-mind reasoning, in adults and children.

CHAPTER II

Study 1. Continuity and Change in Theory-of-Mind Neural Correlates from Early to Middle Childhood: A Longitudinal Study Using EEG and fMRI

Introduction

Having a "Theory of Mind"—the understanding that people's actions are guided by internal mental states such as beliefs, desires, and intentions—is a cornerstone of social-cognitive development (Harris 2006; Wellman, 2002). Theory of mind goes beyond social perception; it requires conceptualization of and reasoning about people's mental states in order to accurately predict and explain behavior.

Over 25 years of research on theory of mind has revealed much about this cognitive phenomenon. There are now abundant studies demonstrating robust behavioral developments in infancy (see e.g., Sodian, 2011 for review) and early childhood (e.g., Wellman, Cross & Watson, 2001; Wellman & Liu, 2004). And neurocognitive investigations on theory of mind in adults (see e.g., Carrington & Bailey 2009; Apperly 2011 for reviews), as well as some recent investigations in children (e.g., Sabbagh, Bowman, Evraire & Ito, 2009; Saxe, Whitfield-Gabrieli, Sholz, & Pelphrey, 2009), reveal a network of neural regions consistently recruited for theory-of-mind reasoning. Yet, the mechanisms and processes underlying theory-of-mind development are still unclear.

Indeed, several alternative accounts of theory of mind and its development exist. At the least, some scholars predict uniform and stable development of innate "theory-of-mind modules" that are in place in infancy or come on-line soon after (e.g., Scholl & Leslie, 1999). Whereas other scholars predict that theory-of-mind understanding builds progressively and variably over development, as children's initial naïve theories about the mind are revised and changed based on new experience (e.g., Gopnik & Wellman, 2012). Despite the wealth of existing research, fundamental questions of both developmental *continuity* and *change* remain unsettled.

Behavioral research frames current understanding of theory-of-mind development. Converging evidence demonstrates that infants' mental-state understanding advances from initial unsteady links between action and intention (e.g., Woodward, 1998; Brandone & Wellman, 2009) to more solid conceptions of person-specific desires (e.g., Phillips, Wellman, & Spelke, 2002; Repacholi & Gopnik, 1997), and perhaps even to an implicit (non-verbal) understanding of false beliefs (e.g., Onishi & Baillargeon, 2005) (though the nature and limits of this implicit understanding are currently debated—e.g., Perner & Ruffman, 2005; Sodian 2011). And numerous studies tracking theory-of-mind developments beyond infancy show that critical advancements in belief-understanding manifest during the preschool years, whereby children transition from consistently failing to consistently passing standard explicit (i.e., verbal) false-belief tasks (Wellman et al., 2001). Moreover, behavioral research demonstrates that this preschool development proceeds in a progression of mental-state understandings; children consistently develop an explicit understanding of desires around ages 2 and 3 years, before developing an explicit understanding of beliefs around ages 5 and 6 years (e.g., Wellman & Liu, 2004).

However, there are limits to what can be learned about theory-of-mind development from behavioral research alone. As one example, children's performance reaches ceiling on the standard behavioral theory-of-mind tasks at roughly age 6 or 7 years (see e.g., Wellman et al, 2001; Wellman & Liu, 2004). Moreover, few behavioral tasks exist that capture belief and false-belief understanding beyond early childhood, and even fewer exist that measure progressions of development for multiple mental-state concepts. And moreover, as noted above, several possible underlying mechanisms can predict or account for the same sequences of development.

Neuroscientific methods provide a promising addition to behavioral studies in service of investigating outstanding questions of developmental continuity and change. Identification of neurological correlates of cognition (neural substrates, networks, and signatures that are functionally related to cognitive processing) provide insight into cognitive development generally, and theory of mind specifically—especially when behavioral measures are limited or ambiguous. Of particular relevance to the present research, neural correlates of theory-of-mind reasoning identified at early points in development can be compared with neural correlates present in older children and adults to investigate differences and similarities pointing to continuity and change. In contrast, behavioral methods spanning this wide age range are typically so discrepant that they obscure developmental comparisons. The purpose of the present study is to use longitudinal electroencephalographic (EEG) and functional magnetic resonance imaging (fMRI) data to examine continuity and change in the neural correlates of theory of mind over early to middle childhood.

Emergent Patterns of Continuity and Change in Neurocognitive Research

Existing neurocognitive investigations of theory of mind already shed some light on issues of developmental continuity and change. Abundant neurocognitive investigations of theory of mind in adults converge on the finding that theory-of-mind reasoning consistently recruits a network of specific neural regions: the medial prefrontal cortex (MPFC) and the left and right temporoparietal junction (TPJ) most consistently, as well as the superior temporal gyrus/sulcus (STG/S), the precuneus, and the temporal poles. These regions are recruited when adults engage in multiple mental- and social-reasoning tasks across fMRI and electrophysiological (EEG/ERP) methods alike (see Apperly, 2011; Carrington & Bailey, 2009; for recent reviews review). Only recently have there been examinations of theory-of-mind neural correlates in children, and the pool of existing pediatric studies is small. However, when adult neural signatures are compared with findings from investigations of theory-of-mind neural correlates in child populations, patterns of neurocognitive continuity and change already begin to emerge.

To illustrate, when considering neurocognitive investigations of theory of mind in children, findings implicate a network of regions remarkably similar to the network identified in adults. For example, in their fMRI study examining action-intention understanding and mental-state attribution in typically-developing children aged 7- to 13-years-old, Ohnishi and colleagues (2004) found activation in the bilateral STS, temporal lobes, and fusiform gyrus in both action-intention and mental-state attribution conditions, as well as unique activation for the mental-state attribution condition in the right MPFC, the right inferior parietal cortex, and the right TPJ. These findings implicate almost all of the same neural regions identified in the adult theory-of-mind neural network (e.g., STS, temporal poles, MPFC, and TPJ). Likewise, Sabbagh, Bowman, Evraire, and Ito (2009) used dense-array (128-channel) EEG recordings to investigate the relation between 4-year-olds' resting EEG alpha coherence (a measure of functional maturation of underlying neurocognitive systems; Nunez, 1995; Thatcher, 1992; 1994) and their performance on a battery of theory-of-mind tasks (e.g., appearance-reality, false-belief). Source localization of the EEG alpha showed that increased functional maturation of the dorsal MPFC and the right TPJ predicted increased theory-of-mind performance in these children (even after statistically controlling for domain general constructs also known to be related to theory-of-mind reasoning such as executive functioning, vocabulary development, and age), providing additional evidence that the TPJ and dorsal MPFC are associated with specifically theory of mind even as young as 4 years old.

Thus there are similarities in neural networks recruited for theory-of-mind reasoning in both adults and children. Note that comparisons between child and adult neural data could have turned out differently; conceivably the adult data could have shown neural activations only for expert, fluent, well-developed social cognition—activations that were the products of child and adolescent developments versus contributors to earlier development. However, given that many of the adult neural regions are also implicated in the development of theory-of-mind reasoning from early on, the data illustrate some important neurodevelopmental continuities.

Yet neuroscientific research also reveals evidence for potentially important change in the neural substrates underlying theory-of-mind reasoning. In their cross-sectional fMRI study, Saxe, Whitfield-Gabrieli, Scholz, and Pelphrey (2009) measured neural activation as typically developing children (6-to 10-years-old) listened to stories in three conditions: 1) a mental condition (descriptions of peoples' mental states including intentions, desires, and beliefs), 2) a social condition (non-mental descriptions of peoples' appearance and interactions), and 3) a physical condition (descriptions of physical scenes). Results showed greater activation in the bilateral TPJ, precuneus, and MPFC for the mental condition relative to the physical condition. Additionally, as children aged, specifically the right TPJ was found to increase in selectivity for mental-state reasoning (in comparison to both processing physical descriptions *and* descriptions of peoples' appearance and interactions), demonstrating change in this posterior neural region and continued development beyond early childhood, beyond the age at which children's performance reaches ceiling on behavioral measures of theory of mind. Indeed, the authors argue that this later development contradicts nativist accounts of theory-of-mind development, which emphasize innate theory-of-mind modules already present at or shortly after birth (e.g., Scholl & Leslie, 1999).

Elements of both continuity and change in theory-of-mind neural correlates are even more apparent in studies that directly compare theory-of-mind reasoning in children and adults. As background, often when comparing child and adult neurocognitive activity, children show greater (and more diffuse) activation compared to adults (Casey, Giedd & Thomas, 2000). This pattern has been demonstrated in investigations of theory of mind specifically. For example, Pfeifer and colleagues (2009) measured fMRI activity as adults and young adolescents (ages 11 to 14 years) engaged in direct and metacognitive self-thought (e.g., "What do I think about myself? I think I am…" / "What does my friend think about me? I think they think I am…"). Both adolescents and adults showed activation in the MPFC, left TPJ, posterior STS, and areas of cingulate cortex, but adolescents showed greater activity in each of these areas overall, compared to adults.

Beyond global differences in magnitude of activation, research also demonstrates differential activation patterns for children compared to adults and even for younger children compared to older children. These differences are exhibited most often in terms of how theory-of-mind reasoning activates the MPFC and TPJ at different ages. For example, in their fMRI study, Moor and colleagues (2012) found differences in magnitude as well as location of activation when adults, early adolescents (ages 10 to 12 years) and mid-adolescents (ages 14 to 16 years) inferred mental states from images of eyes. At all ages, activation was observed in the posterior STS. However only the youngest age group showed additional involvement of the MPFC, the inferior frontal gyrus, and the temporal poles, and regression analyses showed decreasing activation in these areas as children aged.

Sommer et al. (2010) also demonstrated increased activation in the MPFC for children versus adults. In their fMRI study, they had adults and children (ages 10 to 11 years) view cartoons depicting characters' true and false beliefs. Both age groups showed increased activation in the dorsal MPFC and dorsal anterior cingulate cortex for false-belief reasoning compared to true-belief reasoning; however, the activation in the dorsal MPFC was significantly greater in children compared to adults. As an additional adult-child difference: activation was found in the TPJ for false-belief versus true-belief reasoning in adults, but not in children.

An increased role of the TPJ for older versus younger age groups was also demonstrated in Gweon et al. (in press). FMRI activation was measured as adults and children (5 to 11 years of age) listened to descriptions of peoples' mental states (mental condition), peoples' appearance and social interactions (social condition), and physical scenes (physical condition). Both adults and children showed greater activation for the mental condition compared to the physical condition in the left and right TPJ, the dorsal

MPFC, and the precuneus. However, adults showed higher selectivity for specifically mental-state processing—relative to physical *and* social processing—in the TPJ and precuneus, compared to children. Moreover, correlation analyses showed that this selectivity for specifically mental-state reasoning increased with age in both the left and right TPJ, and selectivity in specifically the right TPJ positively correlated with children's behavioral theory-of-mind performance.

Taken together, findings from the above studies demonstrate some emergent patterns of continuity and change with respect to the role of the MPFC and TPJ in theory of mind. Results demonstrate that the TPJ becomes increasingly recruited (Sommer et al., 2010; Saxe et al., 2009; Gweon et al., in press) and increasingly specialized/selective (Saxe et al., Gweon et al.) for mental-state processing as children age from as early as 5-years-old to adulthood. And Gweon et al. shows that this increase in selectivity is also correlated with increased behavioral performance on belief-reasoning tasks. Thus, there is converging evidence for a potentially more specialized role of the TPJ as children get older and improve in their understanding of mental states. Relatedly, there is complementary evidence for continuity in MPFC as recruited for theory-of-mind reasoning across younger ages, as well as potentially a more prominent role of the MPFC in younger children versus older children and adults. The MPFC is recruited for processing thoughts and beliefs more strongly in children (10-14 years) compared to adults (Pfeifer et al., 2009; Sommer et al., 2010). It is involved in inferring mental states for younger children (10-12 years), but not for older children (14-16 years) or adults (Moor et al., 2012). And it is related to theory-of-mind reasoning in 4-year-olds, independent from any common associations with executive functioning (Sabbagh et al., 2009), but does not necessarily show this domain-specificity in adults (Saxe, Schulz & Jiang, 2006).

Although these emergent patterns provide initial, tenuous evidence for both continuity (e.g., in the MPFC from 4 through 14 years) and developmental change (e.g., in the TPJ from 5 years through adulthood, and potentially in the MPFC beyond 14 years) in neural regions that support theory of mind, they involve cross-sectional comparisons between children across age ranges, and between separate studies. The nature and extent of any continuity and change outlined in this way may be masked by differences in methods across studies, and various cohort effects. Investigations that track the neural correlates of theory-of-mind reasoning longitudinally over the developmental time course, in the same sample of children, are therefore vital to obtaining a clearer picture of continuity and change in theory-of-mind development. The present study undertakes such a needed longitudinal investigation.

The Present Study

Four-year-old children from the Sabbagh et al. (2009) study described above (for which we had both source localized EEG alpha data as well as theory-of-mind behavioral performance data) were examined again, three years later, when the children were 7- and 8-years-old, forming the first and second wave of a longitudinal study. At this older age range, I used fMRI to measure the neural correlates of theory of mind directly. Thus, the present study has the distinct advantage of examining the same group of children over time, as a direct measure of continuity and change in the neural correlates of theory of mind from early to middle childhood. Moreover, potentially, the present study provides two sorts of needed methodological information as well. First, it can provide validation for the results and methods of our initial EEG localization study in the form of fMRI data. Second, the direct relation of EEG to fMRI data in the same sample of children allows for cross-methodological validation on a larger scale, demonstrating how electrophysiological data (collected at a lower initial spatial resolution) can predict hemodynamic data (collected at a much higher spatial resolution with fMRI). I briefly outline rationale behind wave 1 and 2 below.

Wave 1. As outlined in Sabbagh et al. (2009), at wave 1, when children were 4-years-old, we measured their performance on a battery of standard behavioral theory-of-mind tasks including measures of false-belief understanding and understanding of appearances versus reality. We also collected baseline/resting EEG data as children viewed a static image. Children's behavioral theory-of-mind performance was then related to their resting EEG activity to determine which neural regions were associated with theory of mind at age 4 years.

Electrophysiological measures provide one of the most widely used and reliable windows on young children's neurocognitive development given the quiet, non-invasive, child-friendly aspects of EEG acquisition. We focused EEG analyses on amplitude (power) in the 6-9Hz "alpha" frequency band in order to index functional maturation of neural regions that might be related to cognitive advancements (i.e., in theory of mind). From infancy through preschool, amplitude (power) in this band gradually becomes the highest amplitude resting rhythm over all regions of the scalp (Marshall, Bar-Haim, & Fox, 2002). Moreover, important regional changes in alpha *coherence* (the extent to which spectral EEG is correlated at any two electrode sites) are also exhibited during this time (Thatcher 1994; Thatcher, Walker, & Guidice, 1987). In general, increases in EEG coherence reflect increases in synchronized neural firing either within or across neural populations (Nunez, 1995). Thus, when applied in the developmental context, increases in resting alpha coherence are generally thought to reflect functional maturational changes in the organization of neurocognitive systems (Thatcher, 1992).

Advances in EEG analysis offer ways to estimate neuronal sources of scalp coherence patterns. In particular, the sLORETA method uses the *cross-spectral matrix* (essentially a matrix of all possible pairwise coherence measures within an EEG recording montage) to estimate the intracerebral sources of spectral EEG power (Pascual-Marqui, 2002). This technique estimates the current density at all points in the solution space (6,239 5mm^3 voxels), and has been validated in event-related potential (ERP) studies (see Pascual-Marqui, Esslen, Kochi, & Lehmann, 2002, for a review) including studies that used group differences in regional current density to predict aspects of cognitive and affective information processing (Pizzagalli, Peccoralo, Davidson, & Cohen, 2006; Pizzagalli, Sherwood, Henriques, & Davidson, 2005). At wave 1, we used this technique to measure individual differences in regional current density of EEG alpha. Given the previously described relations between resting EEG alpha and functional neural development, regional increases in current density estimates (via sLORETA) can be taken to reflect increased synchronous activity within associated neural assemblies, which in turn can be attributed to ongoing neurodevelopmental processes and maturation (Thatcher, 1992).

Therefore, at wave 1, we examined the extent to which individual differences in 4-year-olds' regional functional maturation (based on sLORETA current density estimates of resting EEG alpha power) related to their behavioral theory-of-mind performance in order to investigate whether neuromaturational developments were associated with developments in theory of mind. To that end, we found that increased sLORETA activations in the right TPJ and dorsal MPFC predicted concurrent increases in theory-of-mind performance. That is, the more functionally mature children's right TPJ and dorsal MPFC were, the more advanced their theory-of-mind reasoning was. Predictive correlations were sizeable in the .51 to .59 range. Importantly, this relation held even after statistically controlling for variables also known to affect theory-of-mind development including age, executive functioning performance, and vocabulary development, demonstrating that functional maturation of the right TPJ and dorsal MPFC was associated with specifically theory-of-mind reasoning (see Sabbagh et al., 2009).

Our efforts thus yielded two important types of data that can be used to predict outcomes at wave 2: 1) behavioral theory-of-mind performance, and 2) intracerebral sLORETA current density estimates of EEG alpha (i.e., indices of neuronal functional maturation). Moreover, findings from Sabbagh et al. (2009) implicating the dorsal MPFC and right TPJ identify two target regions that can inform analyses for the present study by helping define the neural regions of interest in which we expect to find effects. These findings also inform specific hypotheses about the extent to which the TPJ and MPFC—which were related to theory of mind at wave 1—might be recruited for specifically theory-of-mind reasoning at wave 2.

Wave 2. Approximately three years from time of test at wave 1, neural correlates of theory-of-mind reasoning were examined again in a subset of children from Sabbagh et al. (2009), when they were 7- and 8-years-old. In contrast to wave 1, neural correlates at wave 2 were measured via fMRI to assess the neural activations associated with active (task-dependent) theory-of-mind reasoning (versus the neuromaturational links between brain and theory-of-mind investigated with 'resting'/task-independent EEG at wave 1). Specifically, fMRI activity was measured as children listened to stories in three different conditions: 1) mental (descriptions of peoples' mental states), 2) social (descriptions of peoples' appearances and actions, with no reference to mental states), and 3) physical (descriptions of physical objects and scenes). This fMRI task was adapted from Saxe et al., (2009), and was deliberately identical to the fMRI task used in Gweon et al. (in press). Thus, wave 2 data have the additional advantage of replicating findings from these two studies, giving further confidence to the neural regions identified as important for theory-of-mind reasoning in middle childhood.

Specific to the present study, the wave 2 data can make two additional and important contributions. First, wave 2 data were collected via fMRI, which yields neuro-spatial information that is more precise (3.3 mm isotropic for the fMRI data versus 6 to 8 cm for raw EEG data and 5 mm^3 for localized EEG data) and more direct—potentially allowing validation of initial EEG results and methods as well as more general cross-methodological validation between EEG and fMRI designs. Second, wave 2 data directly measure the extent to which neural regions are actively recruited for theory-of-mind reasoning—shedding light on if and how task-independent neurodevelopmental *maturation* can predict task-dependent neural *activity*.

The central advantage of the wave 2 fMRI task is that it measures the extent to which different neural regions are *selective* for processing specifically mental information. That is, percent signal change (from baseline) in the mental condition can be calculated as a proportion of the degree to which it exceeds activation in both the physical condition (which should not systematically activate regions associated with theory of mind) and the social condition (which may activate neural regions associated with theory of mind to the extent that theory-of-mind neural regions are responsive to general ‘social’ information describing peoples’ actions rather than solely responsive to the *mental states* governing peoples’ actions). Thus, this task allows not only identification of which neural regions are actively recruited for theory of mind, but also provides a measure of the extent to which those regions are recruited for processing of *specifically* mental states—yielding a more fine grained examination of how neural substrates are involved in theory-of-mind reasoning. Critically, when we examine relations between these wave 2 active selectivity measures and the neuro-maturational and behavioral theory-of-mind measures at wave 1, the present study is able to investigate 1) whether early functional maturation of neural regions associated with theory-of-mind reasoning predicts later neural specialization of these regions for specifically mental-state processing, and 2) whether this later neural mental-state specialization might also be predicted by early behavioral theory-of-mind proficiency. Thus, the present study provides a direct examination of continuity and change in theory-of-mind development over early to middle childhood, investigating how both early brain maturation and early behavior relate to later neural specialization.

Study 1 Hypotheses

As outlined above, wave 1 and 2 data taken together investigate the possibility that functional maturations of theory-of-mind neural regions (i.e., right TPJ and dorsal MPFC) at age 4 years predict the extent to which those same regions are selectively recruited for theory-of-mind reasoning at age 7 years—an investigation that examines, in essence, how early neural developments contribute to later ones. Likewise, Study 1 also investigates the possibility that theory-of-mind behavioral performance at age 4 years predicts the extent to which theory-of-mind neural regions are selectively recruited for mental-state reasoning at age 7 years—an investigation that examines, in essence, how early behavioral/cognitive achievements contribute to later neural organizations.

In light of the emergent developmental patterns from existing adult and child neurocognitive research, I hypothesize that the data will show evidence for continuity in the relation between MPFC and theory of mind over early to middle childhood. Almost all existing neurocognitive investigations of theory of mind that I know of implicate the MPFC as involved in mental-state reasoning from 4 through 14 years of age. As our sample tracks children from 4- through 7-years-old, I expect that regions of the MPFC that were related to theory of mind at age 4 should predict selectivity for mental-state reasoning in similar MPFC regions at age 7. In contrast, the relation between theory of mind and the TPJ may change over this developmental period. Existing research implicates increasing specialization of the TPJ from as early as 5-years-old. Thus, regions of the TPJ that were associated with theory of mind at age 4 years may not predict regions that show selectivity for mental-state reasoning at age 7 years. Clarification of these developmental patterns will lay an important foundation for future research that could investigate factors contributing to this continuity and change. Thus, Study 1 is an important initial step towards uncovering the mechanisms underlying theory-of-mind development.

Methods

Participants

Sixteen typically developing 7- and 8-year-old children (5 males) were recruited to participate in the second wave of this longitudinal study. The 16 children recruited for wave 2 are a subset of a larger sample of 29 children who participated in wave 1, three to four years prior when the children were 4-years-old (see Sabbagh et al., 2009 for more details on the larger sample of 29 children). All participants were right handed, with normal or corrected-to-normal vision. Although no systematic demographic data were collected, participants were from a predominantly European Canadian, middle class and/or military background, reflecting the demographics of the region in southeastern Ontario, Canada from which they were drawn. Parents reported that all participants were born within 2 weeks of their original due date, were developing typically, and had no history of neuropsychological disorder or trauma. At both waves, children gave assent and parents gave written informed consent prior to any data collection. At the end of the sessions for both waves, parents received monetary compensation and children were given small toy prizes.

Four children were excluded from the wave 2 sample: two children were excluded due to excessive motion artifact in their fMRI data, one child did not want to enter the fMRI scanner at all, and one child was excluded due to an external technical issue that resulted in excessive noise in their fMRI data. The final sample for analyses thus consisted of 12 children (4 males). At wave 1, children in this final sample ranged in age from 49 to 59 months (*M* = 53.67, *SD* = 3.52). At wave 2, ages ranged from 91 to 106 months (*M* = 97.88 months, *SD* = 4.55). The time between wave 1 and wave 2 data collection was 35 to 47 months (*M* = 43.64 months, *SD* = 3.49).

Across the two waves, children were tested by the same experimenter, at the same institution. Methodological details for each wave are described below.

Wave 1 (age 4 years): Behavioral Theory-of-mind and Resting EEG

Wave 1 methods were originally published in Sabbagh et al. (2009). Here, I report only brief descriptions of measures relevant to the current study; procedural details can be found in Sabbagh et al.

Behavioral tasks. Behavioral tasks consisted of a theory-of-mind battery, as well as an executive functioning battery and a language measure (used as covariates). Each of these tasks have been used several times in other studies, and are standard assessments of theory-of-mind reasoning, executive-functioning performance, and vocabulary development for the preschool age group. For the theory-of-mind battery, each task included a 'control' question administered after the target question to ensure that children understood and remembered the key details of the task. Children who answered the control question incorrectly were excluded from analyses. For the executive functioning battery, 'warm-up' exercises specific to each task were performed before any target data collection. These warm-ups were part of the task, and designed to familiarize the participant with how the task worked. The experimenter did not proceed with target data collection until the child successfully completed the warm-up exercises. In the event that a participant did not successfully complete warm-up, target data for that child was excluded from analyses.

Theory of mind battery. *False belief: Contents (Gopnik & Astington, 1988).* Children were shown a Smarties box that contained crayons (instead of chocolates).

Children were asked what ‘‘Mickey Mouse’’ (a puppet) would think was inside the Smarties box given that Mickey had never seen inside the box. Children passed the task if they answered that Mickey would think the box contained Smarties/chocolates (score = 0–1).

False belief: Location (Wimmer & Perner, 1983). Children were shown a scenario in which one character's toy was moved from location A to location B while she was not looking. Children were asked where the character would look for her toy. Children passed the task if they answered that she would look for her toy in location A where she left it (score = 0–1).

Appearance–reality (Flavell, Green, & Flavell, 1986). Children were shown a sponge that was painted to look like a rock and asked what they thought it was. After children responded that they thought it was a rock, they were shown that it was a sponge. Children were then asked again what the object *looked* like. Children passed the task if they correctly answered that it looked like a rock (score = 0–1).

Executive functioning battery. *Grass–snow stroop (Carlson & Moses, 2001).* Children were instructed to reverse typical color-associations and point to a green card when the experimenter said ‘‘snow’’ and to a white card when the experimenter said ‘‘grass.’’ The final score was the proportion (percentage) of correct responses over 16 trials. Sometimes children made multiple responses on a single trial, but only their first responses were scored.

Dimensional-change card sort (Zelazo, 2006). Children were instructed to sort cards that varied on two dimensions: (a) color (red and blue) and (b) shape (boats and rabbits). First, children were instructed to sort cards according to shape (i.e., boats in one basket, rabbits in the other). Then, they were asked to switch and sort cards according to color (i.e., red in one basket, blue in the other). Children were scored based on the number of post-switch sorts that clearly demonstrated they were sorting based on the second dimension (i.e., color; range = 0–3).

Less is more (Carlson, Davis, & Leach, 2005). Children were shown two trays: one containing a large amount of candy (i.e., five jelly beans) and one containing a small amount of candy (i.e., two jelly beans). Children were told that when they pointed to a tray, the candy in that tray would go into ‘‘Naughty Monkey’s’’ cup, and they would get the candy in the other tray to put in their cup. Children received a score of 0 if they pointed to the tray with the larger amount, and score of 1 if they pointed directly to the smaller amount. Scores were summed across 16 consecutive trials (range = 0–16).

Language measure. *Peabody Picture Vocabulary Task (PPVT) (Dunn & Dunn, 1981).* Children were shown a series of picture sets. An experimenter said a word that corresponded to one of four pictures on the page, and children were asked to point to the picture of that word. Word difficulty increased with each set. The task ended when participants made seven consecutive errors. Scores were calculated as the item number of the most difficult correctly identified word minus the total number of errors (range: 0 to 108).

Wave 1 behavioral measure for present study. For the present study, the variable of interest was children’s individual theory-of-mind performance scores, residualized for executive-functioning performance, language performance, and age at time of test. As outlined in Sabbagh et al. (2009), as a first step, scores from each individual task in the theory-of-mind battery were summed to create one theory-of-mind aggregate measure per child. The same was done for the standardized executive functioning scores to create one aggregate executive functioning measure per child. Regression analyses demonstrated that the theory-of-mind aggregate was uniquely related to the executive functioning aggregate, language scores, and age (all *p*s < .05). Thus, standardized residuals from the regression predicting theory-of-mind from these three variables were saved, and used as the final measure for behavioral theory-of-mind performance in the present study. Using the residualized theory-of-mind performance scores assures that any relations between theory-of-mind and brain activation represent the neural processes related to specifically theory of mind, and cannot be accounted for by the relations between theory-of-mind reasoning and these other three constructs.

EEG Measures. ***Rocket ship–spiral line video***. EEGs were recorded during passive, "resting" viewing intervals (unlike event-locked ERPs). During EEG recording, participants watched alternating video clips of a still picture of a rocket ship and an animation of a green line that mapped out a spiral (alternating expanding and contracting spirals with each presentation). The rocket ship and spiraling line components were each 30 s in length and were each presented six times, totaling 6 minutes of video. Only EEG recorded during the rocket ship segments was analyzed; however the alternating video allowed us to maximize collection of artifact-free data because we instructed children that they could move a little during the spiraling line segments (thus allowing them small reprieves), but that they had to do their best to stay still while the picture of the rocket ship was on the screen. This method was adopted from previous research collecting resting EEG from young children (Fox et al., 1995).

EEG recording and analysis: Alpha (6-9 Hz) coherence and current source density. Continuous electroencephalographic data were recorded from the scalp using the Geodesic Sensor Net (EGI, Eugene, OR), a network of 128 Ag⁄AgCl electrodes embedded in an elastic geodesic tension structure. Impedance for all electrodes was kept below 30 KΩ, and all recordings were referenced to the vertex (Cz), sampled at 500 Hz, and digitally filtered between 0.01 and 200 Hz (time constant = 1 s). The raw EEG recordings were filtered (60-Hz notch) and the rocket ship segments were extracted for further analyses. These data were then divided into smaller 2-s segments, and subject to a software algorithmic artifact rejection program (Vision Analyzer; Brain Vision GMBH, Gilching, Germany) that combed the data for evidence of artifact (gradient threshold = ±50 lV in 100 ms, amplitude threshold = ±200 lV, global maximum difference threshold = ±300 lV). Manual inspection of 25% of the EEG records confirmed that these criteria reliably identified artifact due to blink, eye movement, and participant movement. Segments that contained artifact were removed. Only participants who contributed at least 25 artifact-free segments of EEG (i.e., 50 s of data) were considered for analysis. Artifact-free segments were then transformed to average reference to ensure accurate source-localization.

Cross-spectral matrices were created for the single-subject average-referenced data from 5.5 to 9.5 Hz, thereby ensuring we captured activity in the preschoolers’ alpha band (i.e., 6–9 Hz). From the cross-spectral matrix, sLORETA software computed three-dimensional distributions of the standardized current density using standardized estimates of the minimum norm inverse solution (see Pascual-Marqui, 2002, for details). The sLORETA transformation results in activation values for 6,239 ‘‘voxels’’ (5 mm^3)

located within cortical gray matter and hippocampus, as defined by the Probability Atlas from the Montreal Neurological Institute, yielding a final reliable measure of the regional current-source activation values in the 6–9 Hz (alpha) band at each voxel for each participant.

Wave 1 EEG measure for present study. Along the same lines as the behavioral measure, the final EEG variable of interest for the present study consisted of children's individual sLORETA activations (an index of functional maturation of the brain; Nunez, 1995; Thatcher, 1992), residualized for age at time of test as well as the number of useable (artefact-free) raw EEG segments contributed by each participant. Regression analyses revealed that sLORETA activations uniquely related to both age and number of contributed EEG segments (all *p*s < .05). Thus, standardized residuals from the regression predicting sLORETA activations from these two variables were saved, and used as the final measure for functional neural maturation in the present study. Under the same rationale as for the behavioral measure, using the residualized sLORETA activation scores assures that any relations between brain maturation and other variables of interest represent relations with actual underlying EEG alpha coherence, and are not simply reflections of external relations with age and overall amount of data.

Wave 2 (age 7/8 years): fMRI

fMRI Stimuli. To measure task-dependent neural activation for theory-of-mind reasoning at wave 2, I used the same stimuli as those used in Gweon et al. (in press). Stimuli consisted of acoustically delivered stories, read by one of three female speakers in child-directed prosody, designed to fit one of three conditions[1]: 1) *Mental* stories which described characters' thoughts, beliefs, and desires, 2) *Social* stories which described characters' social relationships and physical appearance, and 3) *Physical* stories which described physical states and objects in the world. Critically, though both Mental and Social stories contained social information (about peoples' appearances, actions, and interactions), only Mental stories contained specific descriptions of peoples' mental states (e.g., Sam *wanted* to keep the party a secret so he didn't wish Eric a happy birthday. Eric *thought* that Sam forgot about his birthday and was very *upset*). Additionally, Physical stories did not contain either mental-state or social content. Stories were matched across conditions for number of words (*M* = 51.6 words), number of sentences (4.7), duration (20 s), and Flesch Reading Ease Level (*M* = 90.4). See Figure 2.1 for an example story in each condition.

Following the presentation of each story, participants were asked, "Does this come next?" and then they heard a sentence either that did not match the content of the story just presented (non-match trial), or that was a continuation of the story just presented (match trial). Non-matching probe sentences were randomly drawn from other, unrelated stories. The correct response for non-match trials was "no" (e.g., "no, this sentence does not come next") and the correct response for match trials was "yes" (e.g., "yes, this sentence does come next"). Participants made their responses via button press (left button for "yes", right button for "no"). After the probe sentence was heard, a green check mark appeared on the left side of the screen and a red X appeared on the right side of the screen to remind participants to press the left button for "yes" (match) and the right

---

[1] Participants also heard two additional condition types – stories in a foreign language, and music. Neither of these additional conditions was relevant for the purposes of the current study and thus neither was included in analyses.

button for “no” (not a match). These images remained on the screen until the participant made a button response. After making a response, participants received verbal feedback (e.g., “Great job! Get ready for the next one” for a correct response, and “Alright. Here comes another one” for an incorrect response). Half of the probe sentences were match trials (correct response = “yes”), and half were non-match trials (correct response = “no”), counterbalanced within and across runs. This task was designed to verify attention and create a necessity to listen intently to the stories, but also to be relatively easy so as to avoid provoking performance differences across ages. See Figure 2.1 for a schematic of the task.

Stimuli were presented in four functional runs, each of which consisted of 10 experimental (story) blocks (36 s per block; 2 blocks per condition per run) and 3 rest blocks (12 s per block with one at the beginning of the run, one after 5 experimental blocks, and one at the end of the run) for a total of 6.6 min per run. In total, children heard and responded to eight stories in each condition (24 stories overall). All 12 children included in the final sample completed all 4 runs. The order of conditions within a single run was palindromic (e.g., [rest] A B C D E [rest] E D C B A [rest]). Overall, conditions were counterbalanced within and across participants according to 5 pseudo-randomized orders. In each experimental block, participants first heard the story (20 s), followed by “Does this come next?” (1.5 s), the probe sentence (3 s), a pause (6.5 s) during which participants made their response to the probe, and finally the post-response encouragement (5 s) (see Figure 2.1). During the experimental and rest portions of the task, a colorful, abstract visual image, unrelated to the content of the story, was presented on the screen so that children would not be lying in the dark. The image changed every 5 experimental blocks. Stimuli were presented via Matlab using the Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997; Kleiner et al., 2007) running off an Apple Macbook Pro.

Procedure. Procedures involved an initial introductory meeting with experimenters in a child-friendly laboratory on the university campus. One to two weeks later, children came to the fMRI facility for fMRI testing, which began with a practice ‘mock’ scan, was followed by familiarization with the scanner task, and ended with actual scanning for data collection. During the initial meeting, the experimenters showed children and their parents a booklet designed to introduce fMRI scanning procedures and principles using child-friendly language and pictures. Children and parents had the opportunity to ask questions and get to know the experimenters.

On the day of the fMRI scan, children gave verbal assent and parents gave written informed consent. Prior to scanning, participants were accustomed to the scanner environment and trained to lie still in an MRI simulator (mock scanner). During the simulation, children wore sound attenuating headphones and practiced lying still in the mock scanner while watching a movie and listening to the sounds that the scanner makes during anatomical and functional scans (played on a CD player beside the mock scanner at high volume). If children moved during the mock scan, an experimenter tapped the child’s leg gently to call their attention to the movement and help them lay still (this same physical reminder was employed during actual scanning as well to help children remain still for data collection). Practice continued until participants lay still for at least 5 consecutive minutes. After simulation, participants performed several trials of the experimental task to ensure adequate understanding of the task and scanning procedure, and then entered the real scanner for anatomical and functional data collection. Mock scanning and task practice totaled approximately 20-30 minutes. Anatomical and functional scanning totaled 35-40 minutes.

fMRI data collection and analysis. Data were collected on a 3 Tesla Siemens Tim Trio scanner using a 12-channel head coil at the Queen's University MRI Facility in Kingston, Ontario, Canada. A T1-weighted MPRAGE was conducted to obtain an anatomical image (176 sagittal slices, slice thickness = 1.0 mm, TE = 2.2 ms, TR = 1480 ms, flip angle = 9.0 degrees). Functional data were acquired in the axial plane with echo-planar images covering the whole brain at a resolution of 3.3 mm isotropic voxels (32 slices, TE = 30 ms, TR = 1970 ms, flip angle = 77 degrees). The first 4 volumes of each run were excluded from the analysis to ensure steady-state magnetization.

Analytic procedures were the same as those used in Gweon et al. (in press) and adapted from Saxe et al. (2009). Data were analyzed using SPM8 (www.fil.ion.ucl.ac.uk/spm/software/spm8) and custom software written in Matlab. Participant's functional data were realigned, normalized to an EPI template in MNI space, and smoothed using a Gaussian filter of 5 mm (full-width half-maximum). The experiment was analyzed within a general linear model framework using boxcar regressors. The story and response portion of the task were included as separate regressors (6 regressors[2] in total). Participants' movement in all six dimensions was also included in the model as nuisance regressors. Data were high-pass filtered at 128 s to reduce low frequency noise.

---

[2] These 6 regressors reflect the 5 story conditions (the mental, social, and physical focal conditions plus the two additional conditions that were not relevant to the present study) plus the response portion.

Two focal regions of interest (ROIs) were defined based on results of wave 1 data (Sabbagh et al., 2009): dorsal medial prefrontal cortex (DMPFC) and right temporoparietal junction (RTPJ). At wave 1, source localized (via sLORETA) EEG alpha coherence (i.e., functional maturation of neurocognitive systems, Nunez, 1995; Thatcher, 1992) in each of these regions positively related to children's behavioral theory-of-mind performance; thus, examination of fMRI activation in each of these regions at age 7 and 8 years addresses the question of how these two focal regions may continue to be important to theory-of-mind reasoning in middle childhood, and whether there is continuity or change in each of these regions' relation to theory of mind across early childhood development. Based on previous literature, four additional ROIs were defined in individual participants: left temporoparietal junction (LTPJ), precuneus (PC), and middle, and ventral prefrontal cortex (MMPFC, VMPFC). Each of these additional regions has been implicated in theory-of-mind reasoning in middle childhood as well (e.g., Gweon et al., in press; Saxe et al., 2009); thus examination of fMRI activation in these regions is important to capture the larger picture of theory-of-mind neural correlates at ages 7 and 8 years, and allows exploration of the possibility that theory-of-mind neural regions in early childhood may relate to other regions in the theory-of-mind neural network across development.

Following Gweon et al., (in press) all six ROIs were defined using both anatomical location (MNI coordinates identified in previous literature) and functional activation. That is, because there is no purely anatomical definition that would allow precise characterization of each region in each individual brain, we used anatomical MNI coordinates (from Saxe & Kanwisher 2003; Saxe et al., 2009; and Gweon et al., in press) as spatial landmarks, and defined each ROI as the voxels in a 9 mm radius surrounding the anatomical coordinates that showed the diagnostic function of a) clusters of at least 10 voxels (k >10), that were b) significantly more active to the Mental condition versus the Physical condition ($p < .0005$)[3].

For every participant, functional activation was overlaid on the participant's anatomical image to guide selection of ROIs. If there was more than one cluster that met functional criteria within the same anatomical region, the ROI was defined around the peak voxel with the highest t-value (for the Mental > Physical contrast). If, for a given participant, no voxels met functional criteria within the defined anatomical region, the participant was dropped from analysis for that particular ROI.

Neural responses to the Mental, Social, and Physical stories were calculated in each of these 6 ROIs for each participant using percent signal change (PSC) by (1) averaging the raw blood-oxygen-level-dependent (BOLD) response across all voxels within an ROI for each time point in the experiment, (2) calculating the average BOLD response of the ROI in each condition for each time point after the onset of the stimulus, (3) subtracting baseline (average BOLD response of the ROI during fixation) from these values, and (4) dividing this value by the baseline (i.e., PSC = 100*[BOLD response – Baseline]/Baseline). For the purposes of statistical analyses, PSC was then averaged across the time points during which the story was presented (4-22 s after story onset to account for hemodynamic lag) to get a single PSC value in each region for each

---

[3] A similarly high criteria— Mental > Physical differences significant at $p < .0001$—was used in Gweon et al.; I chose a slightly more lenient criterion to account for our smaller sample. This criterion is still conservative enough to avoid inflated type 1 error, but importantly allows us to better capture individual variation in wave 2 neural data by not excluding participants from analyses that exhibit meaningful 'mental > physical' activation that is just slightly reduced compared to other participants whose activation differences meet a more stringent p < .0001 criteria.

participant (Poldrack, 2006). PSC essentially calculates the brain's BOLD response (neural activation) for a given cognitive condition of interest (e.g., listening to mental stories) relative to the BOLD response during a period of 'cognitive rest' (i.e., fixating on an abstract image while performing no specific cognitive task) as a way to measure the degree to which neural regions become active (relative to their 'resting state') during cognitive processing. Thus, PSC can be compared across conditions as a way of assessing the degree to which neural regions are recruited for different types of cognitive processing as reflected in different conditions.

Wave 2 fMRI measure for present study. For the present study, the key variable of interest was children's individual *mental-state selectivity scores*, which were calculated as a proportion of their average PSC values in the Mental, Social, and Physical conditions. I outline how these selectivity scores were calculated, along with their rational and interpretation below.

Given that the ROIs were defined using the Mental > Physical contrast, the final variable of interest for the present study critically focused on the relative response of the independent third condition, the Social stories. Following from Saxe et al. (2009) and Gweon et al., (in press), I calculated 'response selectivity' in each ROI for each participant. Specifically, I determined the degree to which responses in the ROIs were selective for (i.e., showed privileged/amplified recruitment for) mental state information versus general social information by calculating a *selectivity index* to measure the relative difference in PSC between the Mental, Social, and Physical stories: 100*(Mental – Social)/(Mental – Physical). Because the ROIs, by definition, only consist of voxels that showed higher activation for Mental versus lower activation for Physical conditions, the selectivity index uses the relative magnitude of activation to the Social condition to get an unbiased assessment of the degree of recruitment for mental-state reasoning in a given ROI. Specifically, a low selectivity index indicates that activation in an ROI for the Social stories was about as high as its activation to the Mental stories (i.e., mental and social activation are essentially equivalent and both differentiate from the physical control). A high selectivity index indicates that a given ROI's response to the Social stories was about as low as the response to the Physical stories (i.e., mental activation is unique and differentiates from both the social and physical control conditions which are essentially equivalent). Thus, a high selectivity index indicates that mental activation is higher than both Physical and Social activations, as I elaborate still further in the next paragraph.

To interpret high and low selectivity indices, recall that both the Mental and Social conditions contain general social information about peoples' appearances, actions, and interactions, but only the Mental condition contains additional descriptions of peoples' mental states, and the physical condition contains neither mental nor social information. Therefore, equivalent activation for mental and social conditions (low selectivity index) demonstrates that though a given ROI differentiates general social information from non-social information, it does not further privilege processing of specifically peoples' mental states beyond processing of their general social characteristics, and thus demonstrates that the given ROI is *not* selective for mental state information. In contrast, equivalent activation for social and physical conditions (high selectivity index) demonstrates that a given ROI is recruited for processing specifically mental-state information (that is not present in either social or physical conditions), above and beyond any recruitment for the general social information that also accompanies the mental-state descriptions, and thus demonstrates that the given ROI *is* selective for processing mental states.

Selectivity indices were calculated for each individual participant, in each of the 6 ROIs, allowing comparison of mental-state selectivity across individuals, as well as an examination of the extent to which degree of selectivity for mental-states in the 6 ROIs varies as a function of other factors. Saxe et al. (2009) and Gweon et al. (in press) demonstrated that selectivity in the TPJ varies as a function of age, providing important validation for the measure. Focal to the present study was an examination of the degree to which children's mental-state selectivity varied as a function of a) their performance on a battery of behavioral theory-of-mind tasks, and b) their source localized (sLORETA) EEG alpha coherence (an index of functional maturation of neural regions), both measured at wave 1 (age 4 years).

Results

Results are reported in two sections. The first section describes results of wave 2 fMRI selectivity data on their own. These initial preliminary fMRI analyses are important to verify the quality and sensibility of the fMRI data in order to give confidence to results from focal analyses correlating wave 1 data with this wave 2 fMRI selectivity. For preliminary analyses of wave 1 theory-of-mind performance and EEG variables, see Sabbagh et al., 2009. The second larger section reports results of the focal correlations between wave 1 and wave 2 correlations, which are central to the present study.

Preliminary fMRI Analyses

Scanner task performance accuracy. Analysis of behavioral performance on the fMRI scanner task revealed that all children performed with near perfect accuracy across all story conditions. The percent accuracy for the *Mental, Social,* and *Physical* conditions respectively was: $M = 90.0$, $SD = 12.0$; $M = 92.0$, $SD = 9.3$; $M = 95.5$, $SD = 7.9$. Similarly, reaction times across the *Mental, Social,* and *Physical* conditions respectively were: $M = 4.56$, $SD = .37$; $M = 4.67$, $SD = .44$; $M = 4.60$, $SD = .45$. Neither accuracy nor reaction time differed across conditions, nor did accuracy or reaction time relate to any demographic variables (i.e., age and gender) (all ps > .085). Importantly, the high (and undifferentiated) accuracy across conditions confirms that children understood and focused attention on the task. Indeed, the task was designed to be simple enough to achieve high performance accuracy across conditions so that it would be unlikely that any performance differences would produce differences in neural data. Correlation analyses verified that performance accuracy did not related to fMRI selectivity indices in any of the 6 ROIs (all ps > .06).

Selectivity Indices across ROIs. All 12 participants showed activation that met criteria for ROI selection in the DMPFC, RTPJ, PC, and MMPFC. Eleven out of 12 subjects met criteria for LTPJ and VMPFC ROI selection. These results demonstrate that these six regions all showed significantly greater activation to the mental condition versus the physical control condition, evidencing a general specialization for mental-state processing (versus non-mental processing) across these ROIs. This effect is demonstrated in the top panel of Figure 2.2, which shows clearly greater mental activations versus physical activations in all six ROIs. Thus, results add to the growing body of evidence that the MPFC, TPJ, and PC are part of a theory-of-mind neural network that exists in middle childhood, and importantly they replicate results from Saxe et al (2009) and Gweon et al. (in press) (which both found greater mental versus physical activation in the same six ROIs), giving confidence that our small sample of 12 children yields data in line with extant findings.

The focal variables for the wave 2 data were individuals' selectivity index scores (i.e., [PSC_{Mental} - PSC_{Social} / PSC_{Mental} - $\mathrm{PSC}_{Physical}$] x100), which measure the extent to which given ROIs are selective for specifically mental-state processing, over and above processing of social and physical information (recall that a *higher* selectivity score for a given ROI demonstrates that that ROI is *more* selective for mental-state processing—beyond both physical- *and* social-processing). As background, I compared Mental versus Social activation directly. Paired samples t-tests ($\alpha = .05$) comparing activation in the Mental versus Social conditions indicated significantly greater Mental versus Social activation in the RTPJ, LTPJ, PC, MMPFC, and VMPFC (all *p*s < .05), and though the DMPFC did not show significantly greater Mental versus Social activity, activations in the mental condition did appear greater than social activations (see top panel of Figure 2.2). The bottom panel of Figure 2.2 depicts the average selectivity scores for each ROI, which were derived from the proportion of activity across all three conditions. Critically, all ROIs demonstrated some degree of mental-state selectivity, as evidenced by the fact that all ROIs on average had selectivity scores that were greater than zero. On average, mental-state selectivity was greatest in the RTPJ (*M* = 69.69, *SD* = 38.45) and lowest in the DMPFC (*M* = 39.56, *SD* = 47.75), though a repeated measures ANOVA ($\alpha = .05$) revealed no significant differences in selectivity across the 6 ROIs; *F*(5) = .971*,* *p* = .48.

These selectivity index results are also in line with previous fMRI findings using similar and identical tasks (Saxe et al., 2009; Gweon et al., in press), and again, replicate findings from other pediatric neuroscientific investigations that implicate the MPFC, TPJ, and PC in children's theory-of-mind reasoning (e.g., Sommer et al., 2010; Pfeifer et al., 2009; Kobayashi, Glover, & Temple, 2008). Importantly, though each ROI was similarly selective for mental-states, selectivity indices in each ROI varied greatly across participants—range, *SD* for DMPFC: -87 to 93, 49; for RTPJ: 0.5 to 138, 38; for LTPJ: -3 to 137, 41; for PC: -79 to 199, 64; for MMPFC: -29 to 83, 36; and for VMPC: -30 to 155, 53. Thus there is also appropriate variance for focal correlation analyses examining predictive effects of wave 1 data.

Focal Analyses: Correlations Between Wave 1 and Wave 2 Data

Focal analyses concerned relations between wave 1 and wave 2 data. Wave 1 data formed the two predictor variables of interest, taken directly from Sabbagh et al. (2009), and measured when participants were 4-years-old: 1) children's performance on a battery of behavioral theory-of-mind tasks, residualized to control for covarying effects of age, executive functioning performance, and language ability, and 2) children's resting, source-localized (via sLORETA) EEG alpha coherence (an index of functional maturation of neurocognitive systems; Nunez, 1995; Thatcher 1992), residualized for covarying effects of age and the number of useable EEG segments contributed by each participant. Wave 2 data formed the outcome variable of interest, measured when participants were 7- and 8-years-old, and consisted of children's mental-state selectivity scores in the DMPFC, RTPJ, LTPJ, PC, MMPFC, and VMPFC (higher scores = greater mental-state selectivity). Given results of Sabbagh et al. demonstrating positive relations between theory-of-mind behavioral performance and functional maturation of specifically the DMPFC and RTPJ, our central question concerned relations between brain and behavior across wave 1 and wave 2 time points in these two focal ROIs. Thus, key to the present study are results from examinations of 1) whether theory-of-mind behavioral performance at 4-years-old predicts mental-state selectivity in the DMPFC and RTPJ at 7- and 8-years-old, and 2) whether functional brain maturation of the DMPFC and RTPJ at 4-years-old predicts mental-state selectivity in the DMPFC and RTPJ at 7- and 8-years-old. However, to ensure proper examination of theory-of-mind neural correlates across the larger network of regions implicated in theory of mind, fMRI selectivity data from all 6 ROIs were always considered in all analyses, and the analyses with wave 1 EEG alpha data take both a more targeted ROI and more exploratory whole-brain approach.

Visual inspection of the histograms for the fMRI selectivity scores across the six ROIs revealed varying and non-normal distributions. Thus selectivity scores were converted to ranks (1-12) in each ROI to achieve a more stable distribution, and Spearman's correlation tests were used to compare wave 1 and wave 2 data. I first report results of correlations between wave 1 behavioral theory-of-mind performance and wave 2 fMRI mental-state selectivity, followed by results of correlations between wave 1 EEG alpha (sLORETA activations) and wave 2 fMRI mental-state selectivity.

Behavioral theory-of-mind performance (wave 1) and fMRI mental-state selectivity (wave 2). To determine whether children's behavioral theory-of-mind performance at wave 1 predicted wave 2 fMRI selectivity for mental-state reasoning, I conducted 6 separate Spearman's correlations ($\alpha = .05$) comparing theory-of-mind performance scores with selectivity scores in the DMPFC and RTPJ (focal ROIs) as well as the LTPJ, PC, MMPFC, and VMPFC (additional ROIs). Analyses revealed that behavioral theory-of-mind performance at 4 years showed a significant positive relation with mental-state selectivity at age 7 and 8 years in only one ROI: our focal ROI, the DMPFC (r = .66, p = .020); see Figure 2.3a. This result demonstrates that the extent to which children reasoned accurately about others' mental states at 4-years-old predicted the extent to which their DMPFC showed selectivity for specifically mental-state reasoning, three years later, at 7- and 8-years-old—evidencing a clear link between early theory-of-mind proficiency and later specialization of the DMPFC for theory-of-mind reasoning. No other ROI exhibited any significant relations with behavioral theory of mind, including our second focal ROI, the RTPJ (r = -.27, p = .50); see Figure 2.3b. This lack of relation is particularly intriguing given the RTPJ was associated with theory of mind at both wave 1 and wave 2: functional maturation of the RTPJ was associated with increased theory-of-mind performance at 4-years-old (Sabbagh et al., 2009), and the RTPJ showed the highest degree of selectivity for mental-state reasoning on average at 7- and 8-years-old.

Localized EEG alpha coherence (wave 1) and fMRI mental-state selectivity (wave 2). ***Whole-brain analysis.*** As an initial approach, I investigated the extent to which mental-state selectivity at 7- and 8-years-old in each of the 6 ROIs correlated with functional maturation of *any* region of the brain at 4-years-old (indexed by localized sLORETA current density activations estimated from EEG alpha coherence; Pascual-Marqui, 2002). These whole-brain analyses provide an important initial wide lens as a check that the specific regions we define in the EEG data for the ROI analyses (below) are not so narrow as to exclude correlations between later selectivity and early functional maturation of regions that are close to, but not directly within, the defined EEG ROIs. Moreover, these whole-brain analyses address the possibility that later neural selectivity for mental-state reasoning could be related to early functional maturation of other brain regions that support factors known to be related to theory-of-mind development (such as executive functioning and language), but that are not part of the theory-of-mind neural network per se.

Following the analytic approach in Sabbagh et al. (2009), I conducted voxel-wise Spearman correlations between source localized EEG alpha (sLORETA activation values; one value per 6239 voxels) and children's fMRI mental-state selectivity scores (one score per child). Six separate sets of voxel-wise correlations were conducted, one per each ROI (i.e., DMPFC selectivity correlated with sLORETA activation across the whole brain, RTPJ selectivity correlated with sLORETA across the whole brain, and so on for all 6 ROIs). The large number of comparisons in conjunction with our small sample made it difficult to determine a reasonable significance criterion for the whole-brain analyses through Monte Carlo simulations. Thus, I adopted a cluster criterion similar to the criterion developed from original analyses with the sLORETA data in Sabbagh et al: correlations were considered meaningful if at least 20 contiguous sLORETA voxels correlated with selectivity at $p < .05$. Given that the question of interest was whether increases in regional functional maturation predicted increases in mental-state selectivity, only positive correlations were considered; however all tests were two-tailed to be conservative.

Though whole-brain analyses had the potential to reveal correlations between selectivity indices and functional maturation of *any* part of the brain, results revealed that functional maturation of only one neural region within the theory-of-mind neural network yielded significant results. Under the established significance criteria, correlation analyses revealed that EEG sLORETA activations at age 4 years significantly predicted selectivity scores at age 7 in only one ROI—the focal DMPFC—and, only one cluster of sLORETA activations predicted mental-state selectivity in this ROI. Specifically, sLORETA activations in a cluster of 25 voxels in the medial frontal gyrus (peak coordinate 5, 55, 20; Broadmann areas (BA) = 9/10) positively correlated with mental-state selectivity scores in the DMPFC ROI: *rs* = .59 - .71, *ps* < .05 (see Figure 2.4).

Thus, results of the whole-brain analyses provide initial evidence for some continuity in the relation between theory of mind and regions of the medial prefrontal cortex from 4- to 7-years-old. Broadly, the extent to which regions of the MPFC (i.e., the medial frontal gyrus) exhibit functional maturation early in development predicts the extent to which the dorsal MPFC is selective for processing mental states later in development. ROI analyses (below) clarify the extent to which the 4-year-old 'medial frontal gyrus' region and the 7-year-old 'dorsal MPFC' region constitute similar/the same regions of cortex, as a more direct measure of continuity in the relation between MPFC and theory-of-mind over early to middle childhood.

ROI analysis. As a clearer, more direct investigation of continuity and change in theory-of-mind neural correlates over early to middle childhood, I examined the extent to which mental-state selectivity scores in the 6 ROIs at 7-years-old were predicted by functional maturation (EEG sLORETA activations) of *homologous* ROIs at 4-years-old. These ROI analyses provide the most stringent approach to identifying continuity in neural regions over early to middle childhood because they investigate the extent to which functional maturation of neural systems within a given region predicts neural specialization in that *same* region, three years later. As an additional advantage, these ROI analyses offer more power, by limiting the number of voxel-wise comparisons (as only voxels within the defined ROIs are considered), thereby offering increased control of type 1 error rate (Poldrack, 2007).

I used the grand average coordinates of the ROIs identified in the fMRI data (at wave 2), to define homologous ROIs in the EEG data (at wave 1). Specifically, the EEG ROIs consisted of voxels in a 2 cm radius around the grand average coordinates of the six fMRI ROIs (rounded to the nearest coordinate to match the sLORETA spatial resolution of 5 mm^3), to create six ROI homologues: DMPFC and RTPJ (focal ROIs), and LTPJ, PC, MMPFC, and VMPFC (additional ROIs) (see Table 2.1 for list of centroid coordinates and number of voxels in each EEG ROI). The larger radius for the EEG ROIs versus the fMRI ROIs (i.e., 9 mm) was chosen to account for the lower resolution of the sLORETA data in order to achieve the most comparable ROIs across wave 1 and wave 2 data.

Six separate Spearman correlation analyses compared wave 1 EEG sLORETA data to wave 2 fMRI selectivity data across each set of homologous ROIs. For example, DMPFC sLORETA activation at wave 1 was correlated with DMPFC mental-state selectivity scores at wave 2, RTPJ sLORETA activation at wave 1 was correlated with RTPJ mental-state selectivity scores at wave 2, and so on for all six ROIs. As with the whole-brain analysis, correlations between sLORETA activation values (one value per 6239 voxels) and children's mental-state selectivity scores (one score per child) were voxel-wise. As afforded by the more targeted ROI analyses, Monte Carlo simulations were conducted to determine appropriate significance criterion by identifying cluster-size criteria associated with a family-wise alpha of $p < .05$. Specifically, to control for multiple comparisons, 2000 random permutations of the fMRI DMPFC selectivity scores were correlated with the sLORETA activation values. Tracking the resulting significant correlations (at $p < .05$) that occurred in each EEG ROI yielded a cluster criterion associated with a family-wise alpha of .05 that was unique to each of the six EEG ROIs (see Table 2.1). Importantly, when these steps were repeated using random permutations of the fMRI MMPFC selectivity scores, results were identical, giving confidence that the significance criteria established for each EEG ROI were appropriate across each of the six correlations with fMRI selectivity.

Results of these ROI analyses were exactly in line with those of the whole-brain analyses. Under the established significance criteria, results showed that wave 1 sLORETA activation positively related to wave 2 mental-state selectivity in only one ROI: the DMPFC (rs = .59 - .71, ps < .05, k = 25 contiguous voxels, peak voxel coordinates: 5, 55, 20); see Figure 2.5a and b. All voxels in the significant cluster fell within the defined DMPFC EEG ROI[4], which by definition, consisted of the same DMPFC region that exhibited fMRI mental-state selectivity at 7-years. When this significant EEG cluster was compared to the original 4-year-old DMPFC EEG cluster from Sabbagh et al. (2009) that exhibited a positive correlation with theory-of-mind

---

[4] An additional Spearman correlation between DMPFC mental-state selectivity scores and aggregate sLORETA activation values averaged across an even larger cluster of voxels further confirmed the effect. Specifically, sLORETA activations averaged across an ROI with a 1.3 cm radius (chosen to replicate the size of the cluster that showed positive correlations between sLORETA activations and theory-of-mind behavioral performance in Sabbagh et al., 2009) surrounding the same MNI coordinates as before (5, 55, 20—the average coordinate of the DMPFC fMRI ROI) correlated positively with DMPFC selectivity scores: $r = .59$, $p = .045$.

performance at 4 years, there was almost complete overlap: 19/25 voxels or 76% of the present study cluster overlapped with the voxels included in the original cluster from Sabbagh et al. (see Figure 2.5c and d).

Given the overlap between the significant cluster in the present study and the original cluster from Sabbagh et al. (2009), and the fact that the 4-year-old DMPFC region was defined to match the 7-year-old DMPFC region, results demonstrate that functional maturation of this particular region of the DMPFC at age 4 years predicts 1) concurrent theory-of-mind performance, and 2) later selectivity of homologous DMPFC regions for processing specifically mental states. Thus, these results provide direct evidence for continuity in the relation between theory of mind and DMPFC over early to middle childhood: Our data demonstrate that almost precisely the same regions of the DMPFC that are associated with theory of mind at 4 years are also recruited for theory of mind at 7 years, with early functional maturation of these DMPFC regions predicting later DMPFC mental-state selectivity.

It is important to note that, as with correlations between wave 1 theory-of-mind performance and wave 2 mental-state selectivity, our second focal ROI—the RTPJ—again did not yield any significant correlations: there was no region of the brain at wave 1 (in either the whole-brain or ROI analyses) that predicted wave 2 mental-state selectivity in the RTPJ ROI. Furthermore, a targeted analysis of correlations between wave 2 fMRI selectivity scores in the RTPJ ROI and sLORETA activations in the original cluster of voxels that associated with theory-of-mind performance at 4 years (35 voxels around centroid coordinate 55, -55, 30; see Sabbagh et al., 2009) also yielded no significant results. Thus, in contrast to the DMPFC, these analyses demonstrate a distinct lack of evidence for continuity in the relation between RTPJ and theory of mind over early to middle childhood.

Comparing wave 1 behavior and brain predictors. With respect to the DMPFC, the above analyses demonstrate that wave 2 DMPFC selectivity is significantly predicted by both wave 1 theory-of-mind performance and wave 1 DMPFC functional brain maturation (EEG sLORETA activation). To explore whether a given wave 1 variable (brain maturation or behavioral performance) is a stronger predictor of wave 2 selectivity compared to the predictive strength of the other wave 1 variable, I conducted two final Spearman partial correlations: 1) a correlation between DMPFC EEG sLORETA activation (averaged across the 25 voxels in the cluster) and DMPFC fMRI selectivity—while controlling for theory-of-mind performance, and 2) a correlation between theory-of-mind performance and DMPFC fMRI selectivity—while controlling for DMPFC EEG sLORETA activation. Neither correlation was significant: $r = .47$, $p > .05$; $r = .48$, $p > .05$, for correlations (1) and (2), respectively. Thus, though both wave 1 variables significantly predict wave 2 data, neither predictor accounts for a significantly greater portion of the variance compared to the other, suggesting that both early childhood behavioral proficiency and brain maturation are equally important in predicting middle childhood neural mental-state selectivity in the DMPFC.

Discussion

The present study sought to directly investigate continuity and change in the neural correlates of theory-of-mind reasoning over early to middle childhood by examining relations between behavioral and neurological measures of theory of mind over two waves of a longitudinal sample of typically developing children. Specifically, I investigated the extent to which 4-year-olds' functional brain maturation (measured via source localized resting EEG) and theory-of-mind behavioral performance (wave 1 data), predicted their neural selectivity for mental-state processing (measured via fMRI) three years later at 7- and 8-years-old (wave 2 data). Of particular interest was any continuity and change exhibited in the dorsal MPFC and right TPJ—the two neural regions that were found to relate to theory-of-mind reasoning at wave 1 (Sabbagh et al., 2009).

Figure 2.6 summarizes the key foci and key results of the study. Results provided clear evidence for continuity in the relation between theory of mind and the DMPFC over early to middle childhood. Specifically, wave 2 DMPFC mental-state selectivity was significantly predicted by both wave 1 theory-of-mind behavioral performance and wave 1 functional brain maturation of homologous DMPFC regions. Moreover, the DMPFC region that predicted later DMPFC selectivity overlapped almost entirely with the region of the DMPFC that was found to relate to behavioral theory-of-mind performance at 4-years-old (Sabbagh et al., 2009) (see top panel of Figure 2.6).

No other relations were found between wave 1 and wave 2 data, including relations pertaining to the second focal region of interest, the right TPJ. That is, as depicted in the bottom panel of Figure 2.6, though behavioral theory-of-mind performance related to functional maturation of the RTPJ at 4-years-old (Sabbagh et al., wave 1 data) and the RTPJ exhibited the highest degree of selectivity for mental-state processing at 7- and 8-years-old (wave 2 data), neither behavioral theory-of-mind at 4 years nor functional maturation of the RTPJ at 4 years predicted mental-state selectivity in the RTPJ at 7 and 8 years. This *lack* of evidence for any continuity in the relation between RTPJ and theory-of-mind over early to middle childhood provides some indirect evidence for developmental change in the RTPJ. That is, this pattern suggests that the RTPJ may be recruited for theory-of-mind reasoning differently at 7 and 8 years compared to its involvement in theory of mind at age 4 years (resulting in a lack of relation between wave 1 and wave 2)—a difference that could be the product of developmental change over this time period. I discuss findings in each of these focal regions below.

Continuity in the Relation Between Theory of Mind and DMPFC

The present study is the first to directly examine continuity and change in theory-of-mind neural correlates using longitudinal methods. Though other neuroscientific examinations of theory of mind in early and middle childhood cannot provide directly parallel results or comparisons, the extant literature reveals a pattern of results that is generally in line with findings from our data.

Consider our demonstrated DMPFC continuity over early to middle childhood—including the fact that no other region in the theory-of-mind neural network exhibited predictive effects from wave 1 to wave 2. Existing neuroscientific research indicates that, for younger children (e.g., age 8 years) compared to older children (e.g., age 14 years) and adults, the MPFC plays a more prominent and consistent role in theory-of-mind reasoning compared to how other neural regions in the theory-of-mind network (i.e., TPJ, PC, STS) are recruited at these younger ages (see e.g., Kobayashi, Glover, & Temple, 2008; Pfeifer et al., 2009; Moor et al., 2012). In the context of the present study, such consistent, prominent recruitment of the MPFC for theory-of-mind reasoning in children ages 8 to 14 years, considered cross-sectionally across multiple studies, provides loose, indirect evidence for developmental continuity in the relation between theory of mind and MPFC. Such findings are thus consistent with our data demonstrating continuity in the MPFC. Of course, the longitudinal aspect of the present study yields much clearer, direct evidence for continuity, and does so for still younger children from 4- through 8-years-old. Additional longitudinal research investigating theory-of-mind neural correlates from 8 to 14 years and beyond is necessary to confirm whether the patterns exhibited in the extant literature are truly those of developmental continuity and change. Similarly, investigations of theory-of-mind neural correlates in children younger than 4 years are needed to shed light on the extent to which the MPFC might still be featured prominently at even earlier points in development.

Methodological and theoretical contributions. Directly measuring continuity in the relation between theory-of-mind and DMPFC in the present study provides several contributions. Methodologically, the fact that both the 4-year-old behavioral data and the source localized EEG data strongly and positively predicted effects in the same group of children, three years later, gives confidence to the data and analytic approaches taken at wave 1. Though Sabbagh et al.'s (2009) behavioral battery was standard and highly used, its methods for EEG data collection and analysis were more novel. Few studies examine 'resting' EEG data (compared to more standard 'event-related' ERP approaches), and fewer still employ source localization techniques like sLORETA (Pascual-Marqui, 2002).

The current data, however, show that the regions of the DMPFC that related to theory-of-mind behavioral performance, spotlighted at 4 years (from source localized EEG alpha coherence) were almost precisely the same regions that were also recruited for theory of mind at 7 years (as measured by fMRI). Moreover, early functional maturation of these DMPFC regions (measured via task-independent ‘resting’ EEG methods) longitudinally predicted later selectivity in those same regions (measured via task-dependent fMRI). Not only do these longitudinal findings help cross-validate the methods used at both waves, these results also demonstrate that task-independent *maturational* aspects of given neural regions can relate to how those regions are *actively recruited* for cognitive reasoning as measured with task-dependent methods, even across three years of development, within a spatial precision on the order of millimeters. Such clear links between early neuromaturation and later neurocognitive recruitment have theoretical implications for understanding development of neurocognitive systems more generally, and for theory of mind specifically (as discussed below).

Methodologically, these clear links between our more novel task-independent and more standard task-dependent measures endorse the use of source localized resting EEG methods more broadly. Such methods could provide a valuable tool for assessing neural correlates of cognition in populations in which longer, more involved methods (e.g., fMRI and ERP) are less feasible. Indeed, these source localization resting EEG methods—which involve just 6 minutes of quiet data collection and require no active cognitive involvement from participants—could prove very useful for examining cognitive neural correlates in even younger children, infants, and populations with more sever developmental cognitive impairments, thus greatly broadening the scope of our understanding of the neurological processing associated with theory-of-mind reasoning.

In terms of theoretical contributions, because our data are the first to provide longitudinal evidence for continuity in theory-of-mind neural correlates across development, it is worth considering further the breadth of effects that were found. Not only was DMPFC mental-state selectivity predicted by early DMPFC neuromaturation, it was also predicted by early theory-of-mind reasoning—demonstrating two sorts of continuity in the relation between DMPFC and theory of mind. Knowledge of the existence of such a clear effect of continuity in neural correlates of cognition can inform future research to investigate the developmental factors that contribute to the effect.

The robustness and spatial precision of the effect could suggest that one such factor may be biological or genetic in nature. This possibility is particularly intriguing in light of recent investigations of the effect of genetic polymorphisms and neurotransmitters on the development of theory of mind. Specifically, recent research demonstrates that increases in dopamine positively relate to theory-of-mind proficiency in 4- and 5-year-old preschool children (Lackner, Bowman, & Sabbagh, 2010), and polymorphisms of the dopamine D4 receptor gene (DRD4) also affect preschoolers' behavioral theory-of-mind proficiency, with children with the short gene variant outperforming those with the longer variants (Lackner, Sabbagh, Hallinan, Liu, & Holden, 2012). Dopaminergic fibers along the mesocortical pathway directly target the DMPFC, and dopamine is thought to play a role in cell proliferation and cell maintenance in this region (Kalsbeek et al., 1987; Popolo, McCarthy & Bhide, 2004). Thus dopamine may play a role in the development and maintenance of the relation between DMPFC and theory of mind from the preschool years to middle childhood evidenced in our data. Future longitudinal research on theory-of-mind neural correlates should add genetic assessments to investigate the possibility that dopamine and the DRD4 gene relate to the continuity in the DMPFC exhibited as children develop beyond the preschool years.

Lack of continuity in the relation between theory-of-mind and RTPJ

Our finding of a *lack* of continuity in theory-of-mind neural correlates in the RTPJ may also make some important contributions to neuroscientific investigations of theory of mind and to our understanding of theory-of-mind development. Because of our small sample size, lack of findings warrants a cautious interpretation. Nonetheless, the sample for the present study consisted of 12 subjects and this sample size is exactly in line with other published pediatric neuroscientific investigations of theory of mind (e.g., Saxe et al., 2009; Kobayashi et al., 2008). Moreover, the analytic approach of the present study was designed to combat potential power issues. The whole brain analysis comparing wave 1 EEG data and wave 2 fMRI selectivity adopted a relaxed significance criteria, whereas the subsequent ROI analysis offered a more targeted approach with reduced voxel-wise comparisons to gain increased statistical power (Poldrack, 2007). With respect to the RTPJ, even the follow-up analyses targeting the specific RTPJ region that related to theory-of-mind at 4-years-old did not yield positive results. Moreover, the whole brain analyses demonstrated that not one cluster of voxels positively correlated with RTPJ fMRI selectivity, even when *no* statistical threshold was applied. These aspects of the present study give confidence that the null finding with respect to the RTPJ is meaningful, and arguably not merely a result of an undetected effect due to insufficient power.

Importantly, a finding of a lack of continuity in the RTPJ is in line with existing research examining RTPJ neural correlates of theory-of-mind reasoning across the same ages as the children in the present study—indeed, based on that research, a null effect in the RTPJ was hypothesized a priori because the current literature implicates considerable neurodevelopmental *change* in the RTPJ over these years. Specifically, two recent pediatric fMRI investigations used similar methods to the present study, including similar mental, social, and physical stories to assess mental-state processing, and the same mental-state selectivity measure. Saxe et al. (2009) found that children ages 6 to 10 years showed greater activation in the bilateral TPJ, precuneus, and MPFC when processing stories in the mental condition relative to the physical condition. But as children aged, only the right TPJ was found to increase in selectivity for mental-state reasoning (in comparison to *both* processing physical descriptions *and* general social descriptions of peoples' appearance and interactions). Gweon et al. (in press) replicated these results (using an fMRI task identical to the one used in the present study) with a sample of even younger children ages 5 to 11 years. As with Saxe et al., results demonstrated that mental-state selectivity in the TPJ increased with age, but as an important additional finding, Gweon et al. also demonstrated that mental-state selectivity in specifically the right TPJ positively correlated with children's behavioral theory-of-mind performance, suggesting that neural specialization in the right TPJ for specifically mental-state processing might be linked to increasing accuracy for mental-state reasoning.

Thus, results from Saxe et al. (2009) and Gweon et al. (in press), taken together, provide good evidence for neurodevelopmental change in the RTPJ over early to middle childhood, in the form of an increasing specialization for specifically mental-state reasoning—beyond non-mental and general social reasoning—as children age and as theory-of-mind proficiency improves. In the context of the present study, it is possible that this RTPJ neural specialization evidenced from 5- to 11-years-old could account for the present study's findings of a *lack* of continuity in the RTPJ over this same time period. That is, at 4-years-old, the RTPJ could be involved in theory-of-mind reasoning in a more general, less-specialized way, but by 7- and 8-years-old, the region could have undergone considerable reorganization and specialization such that its involvement in theory-of-mind reasoning at older ages would not be predicted by the characteristics of its involvement in theory of mind reasoning in early childhood. Additional longitudinal research on the neural correlates of theory-of-mind reasoning from 4-years-old and beyond, with sampling at shorter developmental intervals, can shed light on how the RTPJ might undergo such specialization, and how the characteristics of the RTPJ at earlier time points might affect later characteristics. Importantly, the addition of behavioral assessments of theory-of-mind reasoning in such a longitudinal study are needed to address the possibility that increasing accuracy and precision in mental-state understanding is closely tied to neural specialization, and to examine how such ties might appear at different developmental increments.

Conclusions

The present study consisted of two waves of longitudinal data that investigated continuity and change in theory-of-mind neural correlates over early to middle childhood using EEG, fMRI, and behavioral measures. Results yielded clear evidence for continuity in the relation between DMPFC and theory of mind: both children's DMPFC neuromaturation (indexed via source localized resting EEG alpha coherence), and their behavioral theory-of-mind performance (both measures assessed at 4-years-old; wave 1) positively predicted DMPFC fMRI mental-state selectivity at 7- and 8-years (wave 2). Results also yielded indirect evidence for change in the relation between theory of mind and RTPJ: though RTPJ was related to theory of mind at both wave 1 and 2, neither EEG data nor behavioral theory-of-mind performance at wave 1 predicted RTPJ fMRI selectivity at wave 2.

These results make important methodological contributions by providing confidence in the findings and methods of wave 1 (originally published in Sabbagh et al., 2009), and by validating a novel measure of neurocognitive correlates (i.e., source localized resting EEG alpha coherence) that can be extended to infant and developmental pathology research to greatly broaden the scope of our understanding of neurocognitive development. Findings also make important contributions to our understanding of theory-of-mind development specifically. They are the first to provide direct evidence for continuity in theory-of-mind neural correlates over early to middle childhood—a finding that lays important foundation for future work to investigate factors contributing to this continuity, perhaps beginning with investigations of the effect of dopamine and genetic variants of the dopamine transporter. Moreover, our findings (in the context of existing research tracking neural specialization of the right TPJ with age and with increasing theory-of-mind proficiency) raise important questions about potential change in the role of the right TPJ as related to theory-of-mind development over early to middle childhood—questions which can be explored in future research.

Table 2.1. Centroid Coordinates, Size (Number of Voxels), and Cluster Criterion for each of the 6 EEG ROIs.

EEG ROI	Centroid MNI Coordinates	Size of ROI (# of voxels)	Cluster Criterion (# of voxels)
DMPFC	5, 55, 30	109	12
RTPJ	55, -55, 25	115	16
LTPJ	-50, -55, 30	125	15
PC	-5, -50, 40	179	30
MMPFC	0, 5, 10	134	18
VMPFC	5, 55, -10	134	18

Note. Cluster criterion indicates the minimum number of contiguous voxels (each significant at $p < .05$) needed within a given ROI to reach a family-wise alpha of .05.

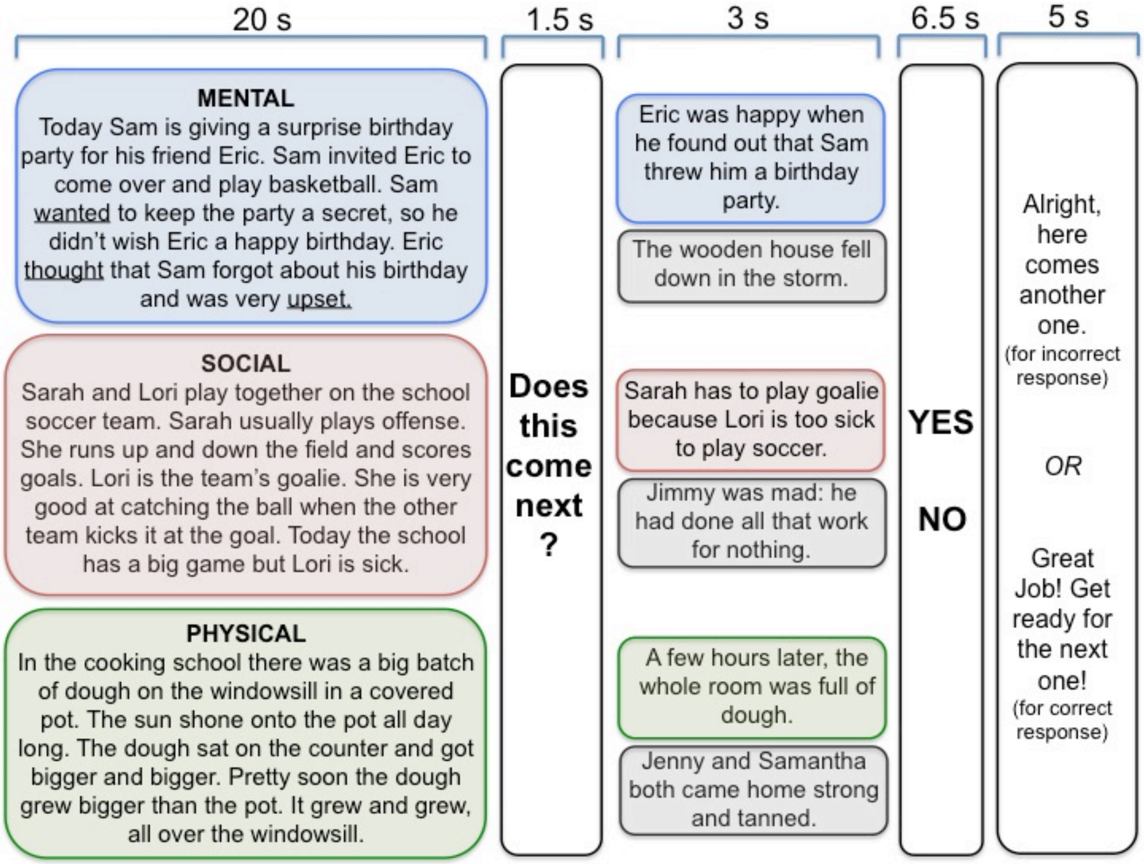


Figure 2.1. Schematic of the task administered at wave 2 for functional MRI data collection (adapted from Gweon et al., in press). The task consists of 5 components outlined in columns from left to right, with durations of each component labeled across the top. Examples of stories in mental (blue) social (red) and physical (green) conditions are shown in the boxes in the first column. The mental-state content unique to the mental condition is underlined. A given story is followed by the question "Does this come next?" (column 2), and then children either here a sentence that continues from the previous story (in matching colors) or a sentence irrelevant to the previous story (in grey). Children judge whether the probe sentence is a match or non-match with the previous story (column 4) and then hear a post-response encouragement that changes depending on whether the participant's response was correct or incorrect (column 5). Only data from the initial story component (column 1) was analyzed; the subsequent task components served to keep children engaged and focused on processing the story information.

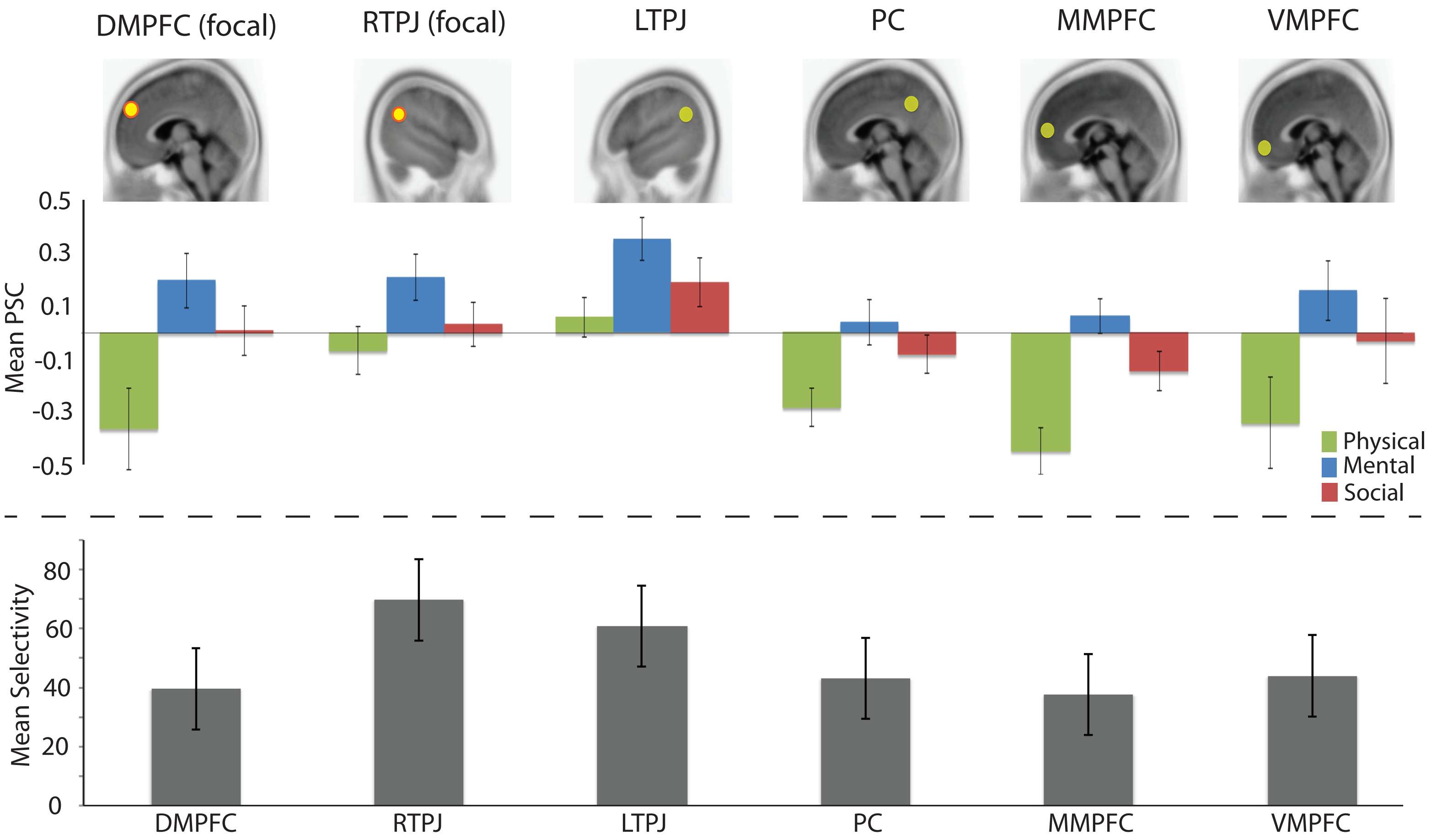


Figure 2.2. Top panel shows percent signal change (from rest) averaged across participants, and averaged across the story blocks (4-22 seconds post story onset) for the physical (green), mental (blue), and social (red) conditions in the 6 regions of interest. All regions show significantly greater mental versus physical activation, as per the functional criteria for ROI selection. All regions also exhibit greater mental versus social activation, suggesting some degree of selectivity for mental-state reasoning across all ROIs (confirmed in bottom panel). Bottom panel shows average mental-state selectivity scores for each ROI (derived from the patterns of PSC depicted in the top panel). This panel shows similar magnitude and variation in selectivity scores across ROIs, with RTPJ exhibiting the highest average mental-state selectivity and DMPFC exhibiting the lowest (though scores did not statistically differ across ROIs).

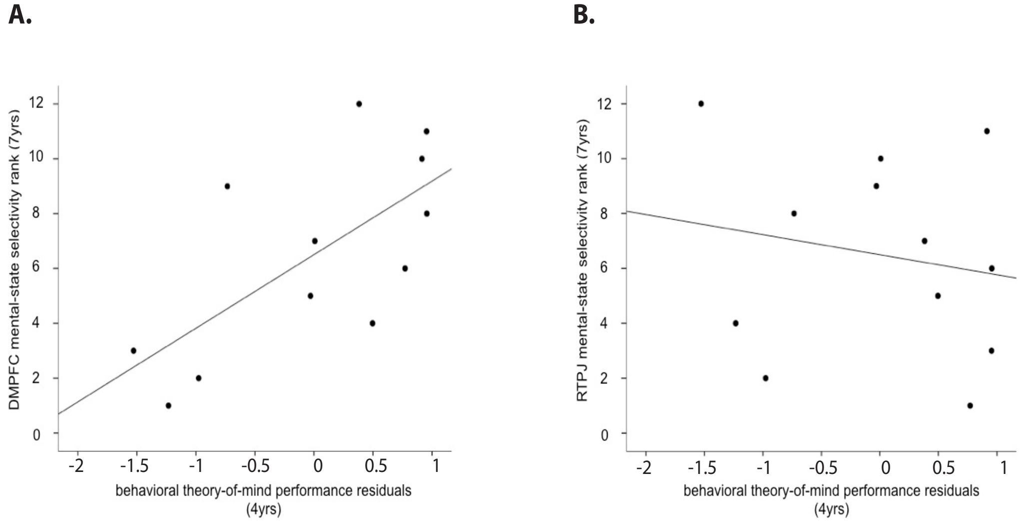


Figure 2.3. Scatterplots of the correlations between wave 1 behavioral theory-of-mind performance (residualized for covarying effects of age, executive functioning and language performance) and wave 2 fMRI mental-state selectivity in the DMPFC (A) and RTPJ (B). The relation between theory of mind performance and DMPFC selectivity was significant. No relation was found between theory of mind and RTPJ selectivity.

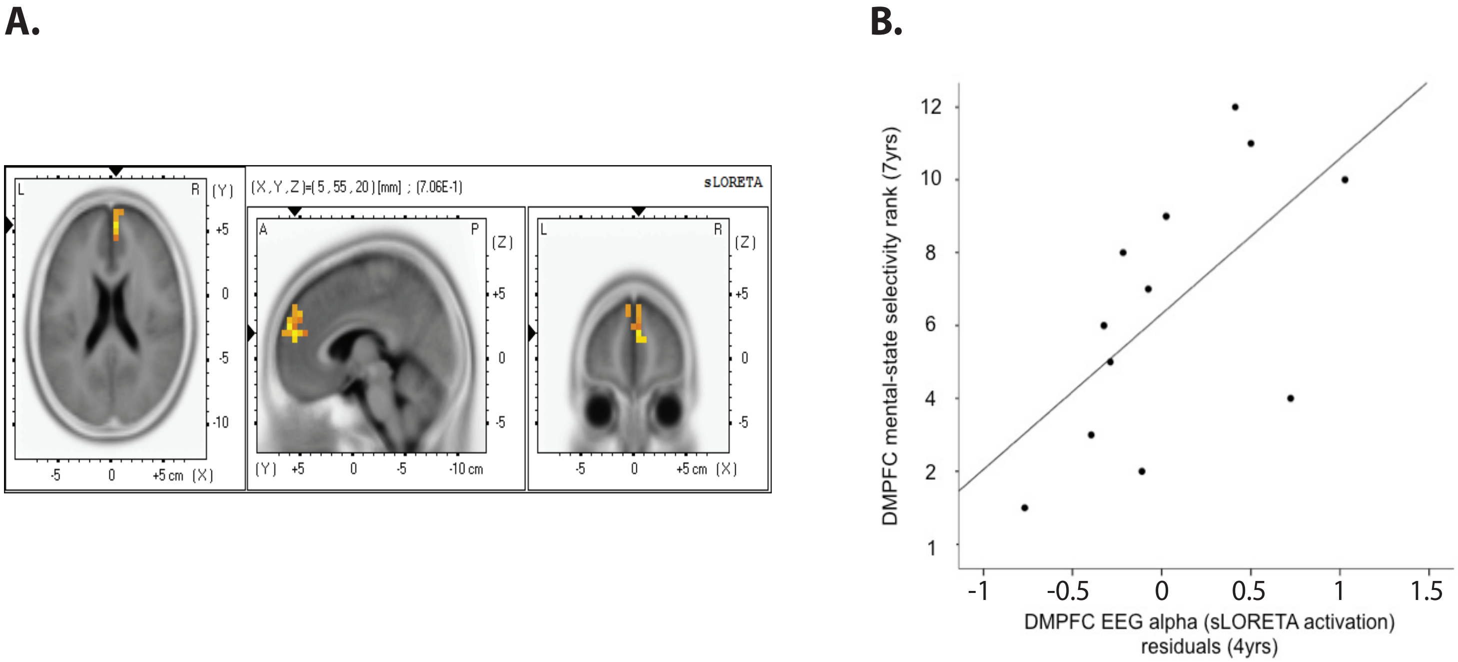
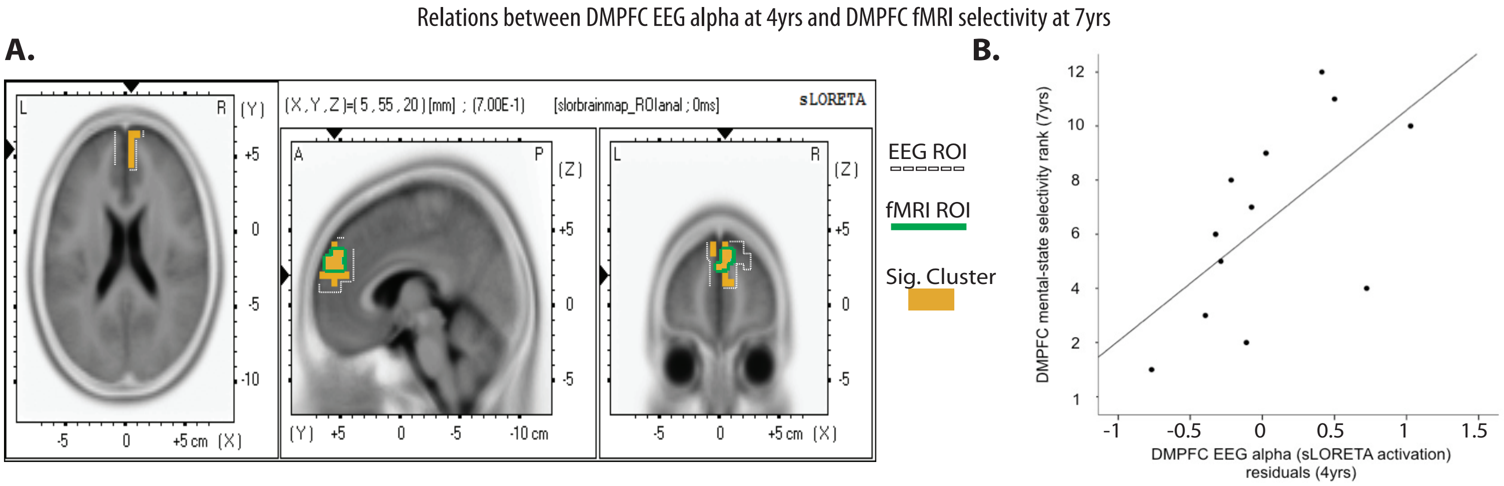
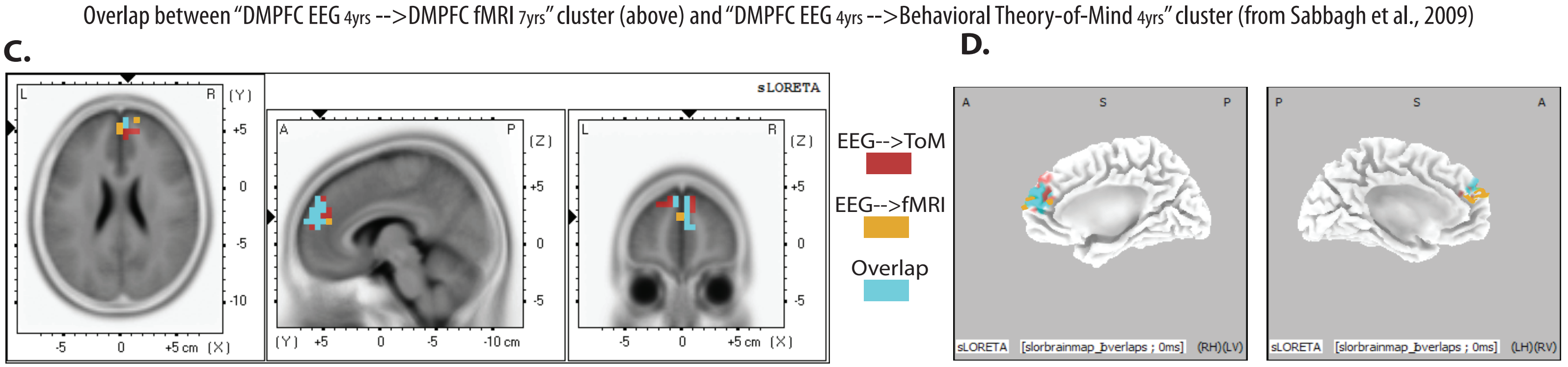


Figure 2.4. Threshold statistical map (A) and scatterplot (B) of the correlations between wave 2 DMPFC fMRI selectivity and wave 1 EEG sLORETA activations resulting form the whole-brain analysis. The significant cluster of EEG voxels that positively related to DMPFC selectivity (at $p < .05$) is shown in orange/yellow (A).



Relations between DMPFC EEG alpha at 4yrs and DMPFC fMRI selectivity at 7yrs



Overlap between "DMPFC EEG 4yrs -->DMPFC fMRI 7yrs" cluster (above) and "DMPFC EEG 4yrs -->Behavioral Theory-of-Mind 4yrs" cluster (from Sabbagh et al., 2009)

Figure 2.5. Top panel: Threshold statistical map (A) and scatterplot (B) of correlations between wave 2 DMPFC fMRI selectivity and wave 1 EEG sLORETA activations, resulting from the ROI analysis. The significant cluster of DMPFC EEG voxels at 4yrs that positively related to DMPFC selectivity at 7yrs (at $p <$ .05) is shown in orange in (A). The 2cm-radius EEG ROI is outlined in dashed white. All voxels in the significant cluster fell within the EEG ROI, and overlapped completely with the DMPFC fMRI ROI (outlined in solid green). Bottom panel: threshold statistical map slice view (C) and alternate cortex map view (D) demonstrating the overlap (blue) between the cluster of EEG voxels that related to DMPFC fMRI selectivity at 7 years (orange), and the cluster of EEG voxels that related to concurrent behavioral theory-of-mind (ToM) performance at 4 years (red). Images depict comprehensive overlap between the two clusters.

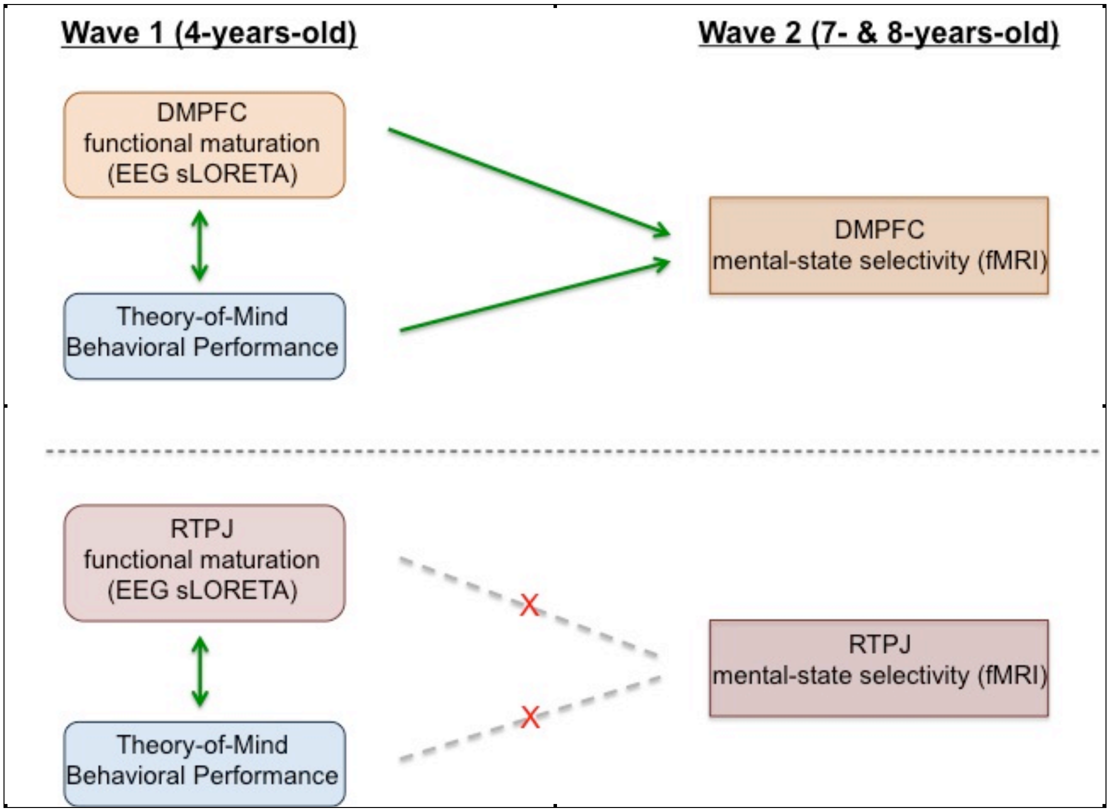


Figure 2.6. Visual summary of correlations between wave 1 and wave 2 data for DMPFC (top panel) and RTPJ (bottom panel). Green arrows indicate significant positive correlations between variables. Grey dashed lines/red x's indicate a lack of correlation between variables.

CHAPTER III

Making Sense of Neurodevelopmental Patterns of Continuity and Change in Theory of Mind: Contributions from Neuroscientific Investigations of Belief- and Desire-Reasoning

As discussed in the previous chapter, Study 1 provides direct and robust evidence for continuity in the DMPFC, and indirect (though corroborated) evidence for change in the RTPJ, over early to middle childhood. But more deeply, what accounts for these developmental patterns? Much is yet to be known about the mechanisms underlying continuity and change, and about the extent of these patterns over the larger developmental course. Several different factors, alone or in combination, could account for the developmental continuity (in the MPFC) and change (in the RTPJ) identified. As discussed in Study 1, genetic components and neurotransmitters, ontogenetic increases and reorganization of grey matter, and increases in performance accuracy and cognitive understanding all offer possible explanations for developmental patterns of continuity and change that should be pursued in future research.

Yet, a key element is missing from Study 1—and indeed is missing from the vast majority of neurocognitive investigations of theory of mind in both adults and children—that would allow for a more complete window on the neurological mechanisms underlying theory-of-mind development, and that would lead to hypotheses of an even broader conceptual and theoretical nature. In general, existing neuroscientific research on theory of mind has focused either on undifferentiated mental-state reasoning (e.g., in Study 1 [and in Saxe al., 2009 and Gweon et al., in press] the “mental” condition consisted of descriptions of peoples’ desires, beliefs, and emotions presented altogether in a single story) or on belief-reasoning alone (e.g., Sommer et al., 2010 examined neural correlates of true-belief versus false-belief reasoning, and Kobayashi et al., 2008 examined false-belief neural correlates across languages). However, as noted at the start, 25 years of behavioral theory-of-mind research demonstrates that different types of mental states (e.g., intentions, desires, beliefs) exhibit different and distinct developmental trajectories. Thus, focus on either undifferentiated mental states generically, or on beliefs alone is limited. Indeed, to reiterate, behavioral data clearly demonstrate that theory-of-mind proceeds in a progression of mental-state understandings, with one of the most robust and highly documented progressions consisting of an early understanding of desires, followed by a later understanding of beliefs. Children exhibit initial explicit desire-understanding around ages 2 and 3 years, but do not show accurate explicit belief-understanding until later in development around ages 4 and 5 years (e.g., Wellman and Liu, 2004). And even in older children and adults, belief-reasoning can be less accurate and fluent (e.g., Malle, 2004). Thus, a direct investigation of the neural correlates for desire-reasoning as compared to the neural correlates for belief-reasoning could be of particular importance in revealing developmental patterns of neural continuity and change.

To illustrate, one recent study has directly examined the neural correlates of belief- *and* desire-reasoning in children. Bowman, Liu, Meltzoff, and Wellman (2012) recorded ERPs when 7- and 8-year-olds performed diverse-desires and diverse-beliefs tasks, as well as physical reasoning tasks as a control condition. The tasks, methods, and EEG acquisition system were identical to those used in the Liu et al. (2009a) ERP study with adults, to enable direct comparisons between child and adult ERP data. Both adults and children showed frontal neural activations for belief- and desire-reasoning that equally differentiated from activations for the physical control, but did not differentiate from each other. In right posterior scalp regions, adults also showed neural activation that was specific to belief-reasoning, distinct from both physical *and* desire-reasoning. This pattern was only partly observed in children. When all child ERP trials were included in analyses (as was done for the adult data), no right posterior belief-desire distinction was evident. Only when analyses were concentrated on the trials in which children judged correctly did right posterior activations for belief-reasoning emerge as distinctly greater than activations for desires. Thus by ages 7 and 8 years, children had already developed neural specializations for reasoning about beliefs and desires that were distinct from the neural activations for physical reasoning—neural patterns that are similar to those found in adults, and located in similar mid-frontal scalp regions. But these same children only partly recruited right posterior regions for reasoning about beliefs as distinct from reasoning about mental-states more generally; children showed activation for belief-reasoning as separate from desire-reasoning *only* when analyses were restricted to correct trials.

Results from Bowman et al. (2012), with its critical addition of examining desire-reasoning as contrasted with belief-reasoning, allow for a more comprehensive and more precise picture of the patterns of continuity and change in theory-of-mind neural correlates exhibited both in the extant literature and directly in Study 1. First, these, results suggest a common neural system for reasoning about *both* beliefs and desires in mid-frontal regions. Accordingly, in the extant literature, the MPFC is consistently implicated in studies (separately) measuring not only beliefs, but also intentions and mental-state reasoning more generally, in children ages 5 through 14 years (e.g., Pfeifer et al., 2009; Sommer et al., 2010; Saxe et al., 2009; Gweon et al., in press), and Study 1 provides direct, longitudinal evidence for continuity in MPFC from 4-years-old to 7- and 8-years-old. It is therefore possible that continuity in the relation between MPFC and theory-of-mind over early to middle childhood exists (at least in part) due to the maintenance of the mid-frontal region of the MPFC as recruited for mental-states generically. Data from Bowman et al. confirm that a pattern of generic recruitment exists at ages 7- and 8-years-old (the same age as the children in wave 2 of Study 1), but future research is needed to investigate whether this type of recruitment might also exist in mid-frontal regions at 4-years-old.

Second, and in contrast, data from Bowman et al. (2012) suggest that what emerges developmentally is a growing specialization of the right posterior brain regions (e.g., right TPJ) for specifically belief-reasoning, and they point to a mechanism whereby developmental increases in accuracy for inferring complex mental states contribute to the development of neural specialization that supports social-cognitive understanding. This pattern is in line with results from both Saxe et al. (2009) and Gweon et al. (in press) indicating a growing specialization for the right TPJ as children age and as theory-of-mind accuracy improves. Critically, data from Bowman et al. further clarify that this specialization may be for belief-reasoning specifically, beyond non-mental reasoning generically, *and* beyond reasoning for other types of mental states such as desires. Such a potential change in the function of right posterior neural regions over early to middle childhood (e.g., from recruitment for mental states more generally—including recruitment for *both* beliefs and desires—to recruitment for beliefs specifically) would account for why in Study 1, the RTPJ data at 4-years-old did not predict RTPJ data at 7- and 8-years-old. Moreover, speculatively, as specialization of the TPJ increases for belief-reasoning specifically, some computational load could be transferred away from the MPFC, potentially contributing to the decrease in MPFC activation, and the increase in selectivity of the TPJ seen in older children and adults (e.g., Moor et al., 2012; Saxe et al., 2006).

Of course, these hypotheses rest on only a handful of studies, and include speculations about the neural correlates of desire-reasoning as contrasted with belief-reasoning in children younger than 7 years, which are currently unknown. There is considerable room, and need, for future research. Importantly though, results from Study 1, and results from Bowman et al. (2012) highlight a need for examinations of the role of accuracy for mental-state reasoning, and a particular need for such investigations to include examination of the neural correlates of different types of mental states—critically *desires* as well as beliefs. Studies 2 and 3 of the dissertation each investigate the neural correlates of belief- and desire-reasoning in child populations, and thus, make further steps towards uncovering the process by which mental-state understandings build to form a complete and expert theory of mind.

CHAPTER IV

Study 2. Neural Correlates of Belief- and Desire-Reasoning in Middle Childhood: Evidence from Functional Near-Infrared Spectroscopy

Introduction

Theory of mind – the understanding that human action is governed by internal mental states such as beliefs, desires, intentions, and emotions – is a complex, cognitive phenomenon. Beyond simple social perception, it requires reasoning about multiple interconnected mental concepts in order to accurately predict and explain behavior. Indeed, to make sense of the simple action of a girl walking to the grocery store, one needs to reason about her underlying mental states: a) she wants some apple juice, and b) she thinks that the juice pitcher at home is empty. Even if in reality the pitcher was full of juice, it is the belief that ultimately guides her action. Likewise, a desire for specifically grape juice would result in the same action if the pitcher were full of apple juice.

For adults, reasoning about the underlying mental states of human action and interaction is an easy and typically efficient process. In contrast, over 25 years of behavioral research on children's theory-of-mind reasoning demonstrates distinct developments over infancy and early childhood (see Sodian, 2011; Wellman, 2012 for recent reviews). An often-studied, clear example of these developments is the achievement of false-belief understanding; for example, understanding that the girl could believe that the pitcher contained apple juice even though it truly contained grape juice.

Numerous studies show that children transition from consistently failing to consistently passing standard false-belief tasks in the preschool and early school years (e.g., Wellman, Cross, & Watson, 2001), demonstrating development of an explicit understanding of beliefs and of the mind more generally.

However, focus on beliefs and false beliefs alone is limited. Theory of mind—often termed a belief-desire or belief-desire-emotion naïve psychology—involves understanding multiple causally interconnected mental concepts, and developmentally, children's theory of mind proceeds in a progression of mental-state understandings. A crucial, well-documented progression is that children consistently develop an explicit understanding of desires before developing an explicit understanding of beliefs (e.g., Bartsch & Wellman, 1995; Gopnik & Slaughter, 1991; Wellman & Liu, 2004). The purpose of the present study is to examine, using functional near-infrared spectroscopy (fNIRS), the neural mechanisms underlying understanding of both desires and beliefs in children—thereby providing insights into how different mental-state understandings build to form a complete theory of mind.

The extant findings of understanding desires before beliefs hold across tasks matched on procedural methodology, linguistic structure, and materials. For example, consider diverse-desires versus diverse-beliefs tasks (used by Wellman & Liu, 2004; and also Wellman & Woolley, 1990). For the diverse-desires task, children are told about a character who *likes* a particular snack opposite to what the child prefers (e.g., likes carrots not cookies). Children are then asked to predict which food the character will choose for snack (the carrot or the cookie). For the matching diverse-beliefs task, children are told about a character who *thinks* his cat is hiding in a location opposite to what the child thinks (e.g., thinks the cat is in the garage but child thinks the cat is in the tree). Children are then asked to predict where the character will look for his cat (in the garage or in the tree). The demands and format for the two tasks are virtually identical except that children must predict behavior based on different mental concepts (desires versus beliefs). Children consistently pass diverse-desires tasks at an earlier age than diverse-beliefs tasks (Wellman & Liu, 2004). Moreover, this desire-belief progression has been demonstrated across different cultures (e.g., Wellman, Fang, Liu, Zhu & Liu, 2006), and it even occurs in populations in which social cognition is impaired or delayed (e.g., in late-signing deaf children and high functioning individuals with autism; Peterson et al., 2005). Recent looking-time studies on infants' false-belief understanding (e.g., Onishi & Baillargeon, 2005) have added fuel to the debate about the age at which children possess the capacity to process others' beliefs, at least implicitly. Nevertheless, studies of explicit performance strongly support a consistent developmental progression of explicit understanding of desires before explicit understanding of beliefs.

Though behavioral research frames current understanding of theory-of-mind development, the mechanisms and processes underlying this development are still unclear. Neuroscientific research can provide additional clarity by identifying underlying neural substrates associated with theory of mind. Neurocognitive studies with adults demonstrate that theory-of-mind reasoning recruits a network of neural regions consistently including the TPJ and MPFC, and often the precuneus, STS, and temporal poles as well (see Carrington & Bailey, 2009; Apperly 2011 for recent reviews). Recent neuroscientific research with children demonstrate that a similar network of regions is also recruited to support theory of mind during middle childhood and adolescence

(Pfeifer, et al. 2009; Mosconi, Mack, McCarthy, & Pelphrey, 2005), as well as in children as young as 4-, 5-, and 6-years-old (Sabbagh, Bowman, Evraire, & Ito, 2009; Saxe, Whitfield-Gabrieli, Sholz, & Pelphrey, 2009; Gweon, Dodell-Feder, Bedny, & Saxe, in press).

Development and the TPJ

Beyond similarities between adults and children, some research also suggests that certain neural regions are recruited for theory-of-mind reasoning differently at different ages. In particular, there is emerging evidence that the TPJ shows specialization (i.e. a more focused and narrowed recruitment of neural substrates for a given type of reasoning specifically) for theory-of-mind reasoning in older children and adults, but that it is less specialized in young children. Considering the adult findings, Saxe and Wexler (2005) demonstrated that adults selectively recruited the right TPJ for processing mental states but not for processing other socially relevant facts about a person (i.e., marital status, family relations, cultural background), and that none of the other regions typically recruited in theory-of-mind reasoning (i.e., medial prefrontal cortex) showed such a specified role. Further, Saxe and Powell (2006) found that adults' activations in the TPJ and posterior cingulate were selectively associated with reasoning about beliefs but not with reasoning about other socially relevant facts such as a person's appearance, or about other specific non-mental internal states such as bodily sensations. Some evidence suggests that adults' *right* TPJ in particular is specific for belief-reasoning, above and beyond any involvement in domain-general computations that are also important for theory of mind such as executive functioning skills (i.e., inhibitory control, attention attribution, working memory). In their adult fMRI study, Saxe, Schulz, and Jiang (2006) found that brain regions recruited during an executive functioning task (bilateral intraparietal sulcus, frontal operculum, middle frontal gyrus, and middle temporal gyrus) did not overlap with any regions recruited for a belief-attribution task (left and right TPJ, MPFC, and anterior STS). A third task that required both executive functioning *and* belief-attribution skills did recruit several of the same neural regions as recruited in the executive functioning task; however, the combined task also selectively recruited the right TPJ, which was *not* recruited in the executive-functioning task. According to the authors, this pattern of results suggests there are distinct, domain-specific cognitive processes for belief-reasoning, and that the right TPJ may be specialized for this domain-specific reasoning.

In children however, the role of the TPJ in theory-of-mind reasoning may not be as specialized. In their fMRI study, Sommer et al. (2010) had adults and children (ages 10 to 11 years) view cartoons depicting characters' true and false beliefs. Both age groups showed increased activation in the dorsal MPFC and dorsal anterior cingulate cortex for false-belief reasoning compared to true-belief reasoning; however, additional activation was found in the TPJ for false-belief versus true-belief reasoning in adults, but not in children. A similar pattern of results was found in Gweon et al. (in press). FMRI activation was measured as adults and children (5 to 11 years of age) listened to descriptions of peoples' mental states (mental condition), peoples' appearance and social interactions (social condition), and physical scenes (physical condition). Both adults and children showed greater activation for the mental (versus physical) condition in the left and right TPJ, the dorsal MPFC, and the precuneus. However, adults, but not children, showed greater selectivity for specifically mental-state processing (relative to physical *and* social processing) in the TPJ. Moreover, correlation analyses showed that mental-state selectivity in both the left and right TPJ increased with age, and mental-state selectivity in the right TPJ (but not the left) positively correlated with children's behavioral theory-of-mind performance. This increase in selectivity for mental-state reasoning in specifically the right TPJ was also demonstrated in Saxe et al. (2009). Using a similar task to Gweon et al., researchers measured neural activation as typically developing children (6-to 10-years-old) listened to mental, social, and physical stories. Results showed greater activation in the bilateral TPJ, precuneus, and MPFC for the mental condition relative to the physical condition. But additionally, as children aged, specifically the right TPJ was found to increase in selectivity for mental-state reasoning (in comparison to both processing physical descriptions and general social processing).

Thus, converging evidence from both adult and child neurocognitive studies demonstrate an increasingly specialized role of the right TPJ for theory-of-mind processing at older versus younger ages, suggesting an intriguing role of right TPJ specialization in theory-of-mind development. However, a key element is missing in each of the above studies that would allow for a more complete window on the neurological mechanisms underlying theory-of-mind development. Specifically, both adult and child neurocognitive studies have narrowly focused on belief-reasoning, or on mental-state reasoning in general. To reiterate, because behavioral data clearly demonstrate distinct developments for different types of mental states, focus on either undifferentiated mental states generically, or on beliefs and false-beliefs alone is limited. In particular, as just noted, theory of mind involves understanding both belief-states—that represent the world—and desire-states—that motivate particular actions within the represented world, to get what one wants—and the most prominent and substantiated developmental theory-of-mind progression in behavioral research is one from robust early desire-understanding to later understanding of beliefs (e.g., Wellman & Liu, 2004; Wellman et al., 2006; Peterson et al., 2005).

Only one recent study has directly examined the neural correlates of belief- and desire-reasoning in children. Bowman, Liu, Meltzoff and Wellman (2012) recorded ERPs when 7- and 8-year-olds performed diverse-desires and diverse-beliefs tasks. As a control, participants performed parallel diverse-physical tasks (requiring reasoning about where different things go). Importantly, the tasks, methods, and EEG acquisition system were identical to those used in a parallel ERP study with adults (see Liu et al., 2009a for original publication of adult data), to enable direct comparisons between child and adult ERP data. In this research, both adults and children showed frontal neural activations for belief- and desire-reasoning that equally differentiated from activations for the physical control, but did not differentiate from each other. In right posterior scalp regions, adults also showed neural activation that was specific to belief-reasoning, distinct from both physical and desire-reasoning. This pattern was only partly observed in children. When all child ERP trials were included in analyses (as was done for the adult data), no right posterior belief-desire distinction was evident. Only when analyses were concentrated on the trials in which children judged correctly did right posterior activations for belief-reasoning emerge as distinctly greater than activations for desires.

Results from Bowman et al. (2012) in conjunction with findings demonstrating developmental specialization of the right TPJ represent a straightforward but important developmental possibility: an understanding of beliefs may build on prior desire-understanding, evidenced by additional substrates in right TPJ regions being recruited for belief reasoning—beyond recruitment for desire-reasoning—as belief-understanding becomes more distinct and accurate. That is, if we consider that the right posterior regions identified in Bowman et al. include the right TPJ, then data from Bowman et al. suggest that the right posterior regions (e.g., right TPJ) identified in the above studies could be showing specialization beyond *general* mental-state reasoning, and could be developing to be recruited for specifically *belief*-reasoning over and above reasoning for other types of mental-states such as desires. Such a pattern has clear implications for behavioral findings demonstrating that children reach an explicit understanding of desires before they come to an explicit understanding of beliefs (e.g., Wellman & Liu, 2004).

However, ERP provides only tenuous evidence for brain localization. Thus, the low spatial resolution of ERP methods leaves unknown whether the "right posterior" findings from Bowman et al. (2012) actually correspond to the right TPJ. In order to clarify how findings from Bowman et al. integrate with findings from other studies implicating specialization of the right TPJ over development, we need to clarify that belief specialization (over and above specialization for desires) does indeed occur in specifically children's right TPJ. Such clarification would shed crucial light on how multiple mental-state understandings may build, developmentally, to form an accurate, expert theory of mind.

The goal of the present study is to provide this needed clarification by measuring 7- and 8-year-olds' neural specialization for belief- and desire-reasoning via fNIRS methods. Participants in the present study span the same age range as participants in Bowman et al., and were given the same diverse-desires, diverse-beliefs, and diverse-physical tasks as used in that ERP study, with the exact same stimuli and test questions. The only differences are in timing and duration of stimuli presentation, which were adjusted to optimize data collection for fNIRS methods. I describe how fNIRS methods facilitate pinpointing of neural activation in particular cortical regions of interest next.

FNIRS Imaging and Methods

In brief, fNIRS uses near-infrared light to detect changes in oxygenated and deoxygenated blood flow as a function of neuronal activity (see e.g., Huppert, Diamond, Franceschini & Boas, 2009 for review of NIRS principles and analyses). Near infrared systems typically emit two wavelengths (around 600 and around 900mm), and this light is absorbed differently depending on the concentration of oxy- and deoxy-hemoglobin in the blood. Thus, light emitter optodes (probes that shine infrared light into the brain) and light detector optodes (probes that detect light scattering back up to the scalp surface) are placed on a participant's scalp, and the differential absorption patterns at each emitter-detector optode pair are used to calculate the hemodynamic response function as an indirect correlate of neural activation. FNIRS can detect changes in the blood-oxygen-level-dependent (BOLD) signal (like fMRI), but can also measure responses of oxygenated and deoxygenated blood species separately (unlike fMRI). Some research demonstrates that patterns in oxygenated blood in particular are a better correlate of neuronal activation (Strangman, Culver, Thompson & Boas, 2002).

FNIRS has been used to measure neural correlates of cognition in adults in both healthy and clinical samples (e.g., Joanette et al., 2008), in infants (see Lloyd-Fox, Blasi & Elwell, 2011 for a recent review), and in children including child populations that span the present study's precise age-range (e.g., Sugiura et al., 2011). Though a relatively new neuroimaging method compared to fMRI and EEG, several advances in fNIRS data acquisition and analysis have been made over the past decade (Boas, Dale & Franceschini, 2004; Huppert et al., 2009), making fNIRS a reliable tool for accessing neural correlates of cognition when one works within the method's limits and capitalizes on its strengths (Aslin, 2012; Lloyd-Fox, 2011).

For the present study, fNIRS methods offer several advantages. Key for our purposes is the ability of fNIRS methods to address the localization issues related to ERP data, and pinpoint neural activation patterns for belief- and desire-reasoning in specifically the right TPJ. FNIRS methods provide higher spatial resolution: 3-4 cm for fNIRS (Huppert et al., 2009) versus 6-8 cm at best for unlocalized EEG data (Ferree, Clay, & Tucker, 2001). Even more importantly, fNIRS methods provide unambiguous localization: optodes are placed on the surface of the scalp and measure changes in oxygenated and deoxygenated blood flow in the surface layers of cortex, 2-3 cm *directly* below the optode (Haeussinger et al., 2011). Structural and functional MRI scans (obtained either simultaneously with fNIRS data collection or separately from a given subject) can provide spatial constraints to better guide optode placement and enhance localization accuracy even further (Boas et al., 2004).

Though localization accuracy is not as high as fMRI, fNIRS methods offer advantages over fMRI. The comparatively low cost, lower susceptibility to movement artifact, and quiet means of data collection that allow children to stay in close proximity to their parents make fNIRS more fitting for collecting data from younger children, especially when administering tasks with large numbers of trials, as in the present study. Moreover, given the goal of the present study—to verify the results from a prior ERP study—an important advantage is that fNIRS, unlike fMRI, offers a testing environment very similar to the environment used with ERP methods. ERP offers the same child-friendly advantages as described above, and both ERP and fNIRS methods involve sensors being placed on and secured to the child's head, requiring restricted head movement. Thus, fNIRS methods provide a testing experience that closely parallels that of the previous ERP study, while offering increased localization accuracy to investigate whether 'right posterior' belief and desire activation patterns correspond to the TPJ. As a final, broad advantage, hemodynamic data collected from fNIRS can be compared with the electrophysiological data from the previous ERP study to provide cross-methodological validation of the beliefs-desires activation effects demonstrated in Bowman et al. (2012).

A drawback of NIRS is that it can record only the surface layers of cortex and cannot penetrate to deeper layers (e.g., Huppert et al., 2009). Fortunately, right TPJ lies in the surface layers of the right posterior cortex. Thus fNIRS is suitable for examining the focal hypotheses for this study, which concern early development and specialization of neural substrates in specifically the right TPJ. In contrast, fNIRS cannot be used to reliably examine MPFC; the MPFC lies deeper within the medial prefrontal portions of the brain, which are not necessarily accessible to NIRS (Haeussinger et al., 2009; Fournier, Combès, Roberts, Braga & Prima, 2011), and in general, the optodes over frontal/forehead regions of the scalp exhibit reduced sensitivity to near-infrared light absorption changes in underlying frontal cortex (Cooper et al., 2012). Thus, we aim to clarify the right posterior findings from Bowman et al. but cannot confirm that the 'mid-frontal' belief-desire activations observed there are specific to the MPFC.

To best understand the task-dependent activations in some focal brain region (e.g., right TPJ) it is important to compare that region to some other region where activation would, by hypothesis, not be expected (Aslin, 2012). These ‘contrast’ recordings are particularly important in ruling out activation patterns due to systemic blood flow changes in the scalp and other tissues that lie above the cortex (Boas et al., 2004). I use recordings from the anterior frontal cortex (AFC) (on the surface layers of the brain), which are not implicated in the theory-of-mind neural network, to provide this needed contrast in the present study. Though these frontal optodes exhibit reduced localization sensitivity, they still provide adequate measurement of neural and scalp hemodynamic responses that can be used to contrast with activation patterns in the focal TPJ regions of interest.

Neurocognitive research on theory of mind in both adults and children has also implicated activity for mental-state reasoning in the left TPJ in addition to the right (e.g., Gweon et al., in press). And additionally, a background hypothesis for much developmental neuroscience research is that activations may be more diffuse, exhibiting less clear hemispheric specialization in children versus in adults (e.g., Casey et al., 2000). Thus, in addition to placing optodes over the right TPJ, I recorded neural activity from optodes over the left TPJ as well, to ensure that any posterior beliefs-desires distinctions were captured. Following from results of Bowman et al. (2012), I expect neural activation in the right TPJ to be greater for belief-reasoning compared to both desire-reasoning and physical-reasoning, providing evidence that the ‘right posterior’ beliefs-desires distinction demonstrated in Bowman et al. includes regions of the right temporoparietal cortex, and further confirming that belief activation in the right TPJ is *greater* than (and not just different from) desire activation. Crucially, this pattern of results would suggest that the specialization for ‘general’ mental-state reasoning in the right TPJ observed over early to middle childhood may actually be due in part to a growing specialization of *beliefs* beyond specialization for other mental states such as desires. Finally, given that the AFC is not implicated in the theory-of-mind neural network, I expect belief- and desire-activation in anterior frontal optodes to be low compared to optodes over the TPJ. This differential anterior frontal pattern would give confidence that the activation patterns observed in the TPJ are products of neural activity in the temporoparietal cortex, not due to systemic noise, or a global activation pattern occurring over the whole head.

Methods

Participants

Twenty-one typically developing children (13 males) ages 6-to 10-years-old participated in the study. All participants were right handed, with normal or corrected-to-normal vision. Performance on standardized verbal and nonverbal intelligence tests from the Kaufman Brief Intelligence Test 2 (KBIT-2; Kaufman & Kaufman, 1990) confirmed this sample had age- and grade-appropriate verbal and nonverbal IQ abilities (verbal IQ: *M* = 29, *SD* = 7.58; nonverbal IQ: *M* =28.64, *SD* = 5.66). Data from the first 4 participants were unusable due to a malfunctioning trigger system that did not properly record timing information; these children were excluded from analyses. Visual inspection of the data for movement artifact revealed that 5 additional participants did not provide enough artifact-free data to yield appropriate hemodynamic response curves, and thus they were also excluded from analyses. One additional child was excluded due to belowchance performance across all conditions (35% accuracy for Beliefs, 30% accuracy for Desires and Physical conditions) suggesting a lack of proper attention to the task. The final sample consisted of 11 children (age range: 74 to 129 months; *M* = 92.9, *SD* = 15.84; 9 males). Importantly, children in this sample were drawn from the same geographic area as that of the Bowman et al. (2012) ERP study and the ages of the two samples do not differ statistically (*t*(28) = -1.03, *p* = .278). Thus, the present study sample reasonably approximates the sample from Bowman et al.

Measures and Procedure

fNIRS tasks. To parallel Bowman et al. (2012) as closely as possible, I used the same tasks as those used in the ERP study: multi-trial diverse-desires, diverse-beliefs, and diverse-physical judgment tasks created to measure participants' neural activity for belief-, desire-, and, as a control, physical-reasoning. The stimuli, test questions, and trial structure were identical to those used in Bowman et al.; only the duration of the trial phases and the overall block structure were changed to optimally capture the hemodynamic response measured by fNIRS. In each trial, participants first received information about two characters with different desires for food/toys (Desires condition), two characters with different beliefs about food/toys (Beliefs condition), or two locations to put food/toys away (Physical condition). A pre-recorded female voice aurally presented information that accompanied pictures appearing on a computer screen. As an example of the information phase for one trial, a child would hear "the boy really likes grapes" as a picture of a cartoon boy appeared at the top of the screen with a picture of real-life grapes below it. Those pictures would then disappear and the child would hear "but the girl really likes celery" as a picture of a cartoon girl appeared on the top of the screen with a picture of real-life celery below it. This information phase lasted 8 s. At the end of the trial, participants heard the target question (details provided below for each condition) and were presented with a picture of one of the two food/toys (e.g., celery) for 1.5 s. A picture of the boy and girl then appeared side-by-side on screen (left and right positions randomized across trials), and participants answered the target question by selecting the boy or girl via button press (left button to select character on left side of screen, right button to select character on right side of screen).

Desire reasoning. For the Desires condition, in each trial, the participant heard about a boy who likes a particular food/toy (e.g., grapes/markers), a girl who likes a different food/toy (e.g., celery/blocks), and a closed box said to contain either a snack or a toy (with boy or girl first being counter-balanced). At the end of the trial, participants heard one of four target questions about what would happen when the box was opened: "Who says 'I'll have some' when they see this?" / "Who says 'I won't have any' when they see this?" (when the story was about food), or "Who says 'I'll play with it' when they see this?" / "Who says 'I won't play with it' when they see this?" (when the story was about toys). The question presented (either positive or negative wording) was randomized in each trial. After the target question, participants were immediately presented with a picture of one of the two foods/toys (e.g., celery/blocks). Note that in this and other conditions, participants were not able to answer the target question until presented with a picture of the food/toy that was in the closed box. After 1.5 s of seeing the revealed food/toy, the boy and girl appeared on screen again and participants answered by choosing one of the two characters.

Belief reasoning. The Beliefs condition followed the same presentation format (including counterbalancing and randomization) as just described, except the content of the information and target questions were about beliefs. Participants heard about a mystery box containing food/toys for a guessing game, and participants were told about a girl who thinks the box contains a particular food/toy and a boy who thinks the box contains a different food/toy. The target questions for this condition were: “Who says ‘I was right’ when they see this?” or “Who says ‘I was wrong’ when they see this?” After the target question, participants were presented with a picture of one of the two foods/toys, and then answered by choosing one of the two characters.

Physical reasoning. The Physical condition provided a non-mental control condition (again following the same presentation format as the other two mental-state conditions). Participants heard about a closed box containing food/toys to put away, and participants were told that the red bin should receive a particular food/toy while the blue bin should receive a different food/toy. Target questions for this condition were: “Where do you put this?” or “Where do you not put this?” After the target question, participants were presented with a picture of one of the two foods/toys that was in the closed box, and answered by choosing one of the two bins.

Commonalities across conditions. Because trials for all three conditions were constructed to have the same perceptual and linguistic structure, any differences between conditions would point to the *mental-state processing* beyond these perceptual and task similarities.

The task was presented in 3 runs of 10 experimental blocks (26 s per block) and 9 rest blocks (15 seconds per block); these durations are within the range of other block designs capturing hemodynamic response in children with both fMRI (e.g., Gweon et al., in press) and fNIRS (Minagawa-Kawai et al., 2011). Experimental and rest blocks alternated, with the experimental block beginning and ending each run. Each experimental block consisted of two trials (13 s per trial; 8 s for information phase plus 1.5 s for target image duration, plus additional 3.5 seconds accounted for by duration of target questions and participant response), and trials within a block were always of the same condition (e.g., two Physical trials or two Beliefs trials). The number of blocks for each condition was distributed as equally as possible across the runs given the odd numbers: each condition type was represented by 3 blocks in two of the runs, and by 4 blocks in one of the runs.

Order of condition blocks was randomized with the stipulation that no one condition type repeated successively. Across condition blocks, half of the trials were about food and half were about toys, and half of the food/toy trials had positively phrased test questions and half had test questions phrased in the negative. Order of phrasing (positive/negative) and trial type (food/toy) was randomized across trials, and balanced across runs. At the end of each run (6.6 min in duration), participants could take a small break to rest or eat some snacks. No data were collected during breaks but the fNIRS optodes were not removed. When all three runs were completed, total experiment time including breaks was approximately 25-30 min. See Figure 4.1 for schematic summary of the tasks.

All stimuli in the experiment were presented using MATLAB (MathWorks), Psychtoolbox Version 3 (developed by Mario Kleiner and colleagues), presented with an iMac “Core 2 Duo” 3.06 (2009 model) with a 27-inch screen, and auditory stimuli were played via the built-in Mac stereo speakers. Parents gave informed consent and children gave written and verbal assent prior to beginning the experiment. Participants were instructed on how to do the fNIRS task immediately prior to data acquisition, and completed a practice task that went through an example trial from each of the three conditions before beginning the actual experiment.

Behavioral Tasks. Standardized verbal and nonverbal intelligence tests from the KBIT-2 (Kaufman & Kaufman, 1990) were administered after the NIRS imaging. Performance on these measures was used to identify any participants who might be performing below the expected cognitive level for their age, and to ensure that all participants included in analyses had enough verbal ability to properly comprehend the fNIRS task. All participants performed at age- and grade-appropriate levels; therefore no exclusions were made based on behavioral performance.

fNIRS Imaging

Data acquisition and procedure. Hemodynamic response was recorded using a Hitachi ETG-4000 with 48 channels acquiring data at 10 Hz (though only 36 channels were used for the present study). The near infrared lasers (emitter optodes) were factory set to 690 nm and 830 nm. Optodes were segregated into three 3x3 arrays each containing 5 emitters and 4 detectors to create 12 channels per array (a channel is defined as the curve of near-infrared light travelling between the emitter and detector from which the hemodynamic response is measured) (see Figure 4.2). Optode separation was 3 cm.

Once the participant was comfortably seated, one array was placed over the participant's forehead (center array), and the other two arrays were placed on either side of the head (left and right arrays). Positioning of the arrays was dictated by the 10-20 system. For the center array, the center optode in the most ventral row was placed over coordinate Fp, and the center optode in the most dorsal row was placed in line with coordinate Fz. For the left and right arrays, the most ventral anterior optode was placed 3 cm posterior to coordinate T3 (for left side) and T4 (for right side), and the most dorsal anterior optode was positioned so as to be 3 cm posterior to the T3-C3 line (for left side) and T4-C4 line (for right side) (see Figure 4.2). Arrays were secured to the participant's head using custom-made fabric ties. These initial placements were designed to maximally overlay the AFC (center array), and the left and right TPJ (left and right arrays). MRI anatomical scanning was used to later pinpoint the particular channels that penetrated these regions of interest as described below. Photographs were taken of the secured optode positions before and after the fNIRS experimental task to ensure that the optodes did not move over the course of the experiment.

ROI identification: MRI coregistration. Separate from the fNIRS imaging and behavioral testing session, an MRI anatomical scan was collected from one child representative of the sample (typically developing male, age: 102 months). Three 3x3 arrays of vitamin E tablets were constructed to exactly mimic the optode arrays (i.e., tablets arranged in same geometrical structure and situated 3 cm apart). These arrays were positioned on the child's head using the same 10-20 coordinates as used to position the optodes (e.g., landmarks at Fp and Fz for center array, and T3/T4 and C3/C4 for the left and right arrays), and were secured in place using MRI-safe medical fabric tape that was wrapped over the arrays around the child's head. With the vitamin E tablets positioned on the child's head, a T1-weighted Fast Spoiled Gradient Echo scan was conducted to obtain the high resolution MRI anatomical image (43 sagittal slices, slice thickness = 3 mm, TE = 5.7 ms, TR = 250 ms, flip angle = 90 degrees, bandwidth = 15.63, F.O.V. = 22 cm). Data were collected on a 3 Tesla General Electric scanner using a quad channel send and receive head coil at the University of Michigan fMRI laboratory facility in Ann Arbor, Michigan, USA. During data collection, the child watched a short, silent cartoon video.

The anatomical image was used to identify which particular fNIRS channels overlaid the left TPJ (LTPJ) and right TPJ (RTPJ) focal regions of interest in our child sample. MNI coordinates for LTPJ and RTPJ were taken from a separate fMRI study that examined theory-of-mind reasoning in 12 comparably-aged typically developing children, 7- and 8-years-old (Bowman, Dodell-Feder, Christopher, Saxe, & Sabbagh, in prep; see Study 1 of this manuscript). This fMRI study defined LTPJ and RTPJ regions of interest based on a combination of anatomical information (from separate localizer tasks as used in previous literature, e.g., Saxe et al., 2009; Saxe & Kanwisher, 2003) and functional activation from a 'mental-state condition > physical control condition' contrast—an approach to ROI definition that has been validated in previous pediatric fMRI examinations that used similar theory-of-mind tasks (Saxe et al., 2009; Gweon et al., in press). The grand average and standard deviation for the left and right TPJ ROIs were calculated for this separate sample of 12 children: -54, -52, 26 +/-6 for LTPJ and 54, -52, 26 +/-6 for RTPJ. For the present study, these ranges of coordinates were used to define reasonably broad left and right TPJ ROIs in the single child's MRI anatomical scan, in order to identify which particular fNIRS channels were positioned over the left and right TPJ. Specifically, via MRIcro software (http://www.mccauslandcenter.sc.edu/mricro/mricro/mricro.html), I used the range of

MNI coordinates to traced out the LTPJ and RTPJ ROI boundaries directly on to the child's MRI anatomical scan. Visual inspection of the scan revealed that vitamin E probes corresponding to fNIRS channels 13 and 16 on the left, and channels 2 and 4 on the right were positioned over top of LTPJ and RTPJ, respectively. Thus data from channels 13 and 16, and from channels 2 and 4, were taken as pinpointing the left and right TPJ regions of interest and were examined in final analyses (see Figure 4.3). FNIRS data from center channel 8 (over anterior frontal regions) were used in final analyses as a contrast/control for the focal TPJ regions of interest.

fNIRS Data Processing and Analysis. After the recording session, fNIRS data were exported and analyzed using custom software in Matlab (The MathWorks Inc.) validated in previous fNIRS studies (e.g., Kovelman et al., 2009; Kovelman et al., 2012), and in line with diffuse optical imaging principles (Boas et al., 2004). Conversion of the raw fNIRS data to hemoglobin values was accomplished in two steps: 1) Assuming constant scattering over the path length, attenuation for each wavelength was calculated by comparing the optical density of light intensity during the task to the calculated baseline of the signal (during rest), and 2) these attenuation values along with sampled time points were used to solve the modified Beer-Lambert equation to convert the wavelength data to a meaningful oxygenated and deoxygenated hemoglobin response (measured as a percent signal change from the rest/baseline portions of the task).

Raw time course data (all channels, all conditions) for each participant were plotted in Matlab and visually examined for motion spike artifacts and signal quality in individual channels. Portions of data in which signal change occurred over a period of time that was too fast to be physiological (i.e., a change in magnitude of response > .2

and occurring in less than 3 s) were removed from analyses. Likewise, individual channels that showed activation at either > 5 or 0 were also removed. After artifact rejection, oxy- and deoxy-hemoglobin responses were averaged across experimental blocks and plotted across time for each individual participant to identify any participants exhibiting an abnormal hemodynamic response (i.e., deviation from the expected response of gradual increase in oxy-hemoglobin and simultaneous gradual decrease in deoxy-hemoglobin over the course of the block as defined in e.g., Boas et al., 2004. See Figure 4.4 for example of appropriate oxy- and deoxy-hemoglobin response curves). Coders visually inspected these response curves in two randomly selected non-focal channels (i.e., channels other than focal channels 2/4 on the right and 13/16) from the left and right arrays to look for general proper task activation (e.g., due to visual, auditory, and general cognitive processing). This visual inspection revealed that general task activation (i.e., canonical hemodynamic response curves) was absent in these non-focal channels from 5 participants, indicating overwhelming artifact in their data, and thus these children were excluded from final analyses (as noted above in 'Participants' subsection). Primarily one coder performed visual inspection of the raw data and hemodynamic response curves, with 20% of the data inspected by two additional coders to ensure reliable artifact rejection.

Visual inspection of the individual hemodynamic response plots revealed some heterogeneity across participants. However, for the left and right arrays, consistently across the 11 focal participants was a clear rise, sustained activation, and fall in oxy-hemoglobin response (and accompanying decrease, sustained decrease, and rise in deoxy-hemoglobin) from 0 s (start of experimental block) to 22 s from block onset. A similar pattern of oxy- and deoxy-hemoglobin response was found in the center array albeit from 4 s to 12 s post experimental block onset (a timing difference consistent with the fact that different brain regions exhibit hemodynamic responses at different rates). Thus, I used these epochs (0-22 s for left/right arrays, 4-12 s for center array) to extract mean oxy-hemoglobin values for final analyses. Simultaneous fNIRS-fMRI recordings demonstrate oxy-hemoglobin is the strongest correlate of the fMRI activation (BOLD) response and best approximation of neuronal activity (Strangman et al., 2001).

To summarize, within each epoch, percent signal change in oxy-hemoglobin values were averaged across blocks of the same condition type, across all participants in the final sample, to obtain a grand average mean oxy-hemoglobin response for each of the three conditions. This grand mean oxy-hemoglobin response was compared across Beliefs, Desires, and Physical conditions to examine the pattern of neural activation associated with each condition in our sample of 6- through 10-year-olds in the two focal regions of interest—right channels 2 and 4 (RTPJ), and left channels 13 and 16 (LTPJ)—and in the contrast control center channel 8 (AFC).

Results

To examine differences in grand mean oxy-hemoglobin response across conditions, I conducted a series of paired-samples *t*-tests contrasting oxy-hemoglobin activations (percent signal change from rest) in Beliefs versus Desires conditions, Beliefs versus Physical conditions, and Desires versus Physical conditions. As directed by the MRI anatomical scan, I examined these condition differences in focal channels 2 and 4 (that represent the right TPJ ROI) and channels 13 and 16 (that represent the left TPJ

ROI), as well as in control channel 8 over the anterior frontal contrast region. I first report behavioral performance results from the fNIRS task, and then report results from the analyses of the neural data, which constitute the main focus of the study.

fNIRS Task Performance Accuracy

As expected, children were better at solving diverse-desires (90.3% correct) and physical control tasks (87.9%) compared to diverse-beliefs (62.3%); Beliefs condition versus Desires and Physical conditions, $t(10) = -6.07$, $p < .001$, $t(10) = -6.93$, $p < .001$, respectively. Desires and Physical conditions did not differ from each other, $t(10) = .91$, $p = .385$. This pattern is identical to behavioral performance accuracy on the parallel ERP task in Bowman et al. (2012), and is consistent with numerous findings demonstrating that, compared to desire-understanding, belief-understanding emerges later in early childhood (see meta-analysis in Wellman & Liu, 2004), and can be less accurate/fluent in older children and adults (e.g., Malle, 2004).

In Bowman et al. (2012), the authors dealt with children's varying accuracy across conditions by examining both the full set of trials in the ERP task as well as only the ERP trials for which children performed correctly. For the present study, the reduced temporal resolution of hemodynamic data and the block design of the fNIRS tasks with two trials per block do not allow for reliable separation of correct and incorrect trials. However, fNIRS data offer clearer measures of individual variation in neural responses in terms of variance in mean oxy-hemoglobin activation across individual participants, as compared to wave-form amplitude in ERP data. This clearer measure of individual variance potentially allows for examination of correlations between performance accuracy and neural activation across conditions to directly investigate the possibility that a growing distinction between belief- and desire-activations in the TPJ may be related to an increasing accuracy for mental-state reasoning. Such correlation analyses are not sensible or reliable with the small sample size for the present study; however I discuss the issue of varying performance accuracy in the discussion section.

Neural Activation Patterns Across Conditions

To thoroughly examine children's neural activation patterns in the focal LTPJ and RTPJ ROIs, I conducted three separate ROI analyses: a group ROI analysis, and two individual ROI analyses each with different ROI inclusion criterion. Each analysis has a slightly different set of assumptions and strengths, and together they reveal a detailed and full picture of the differences in neural activation across conditions. I report results from each analysis in the three sections below. Given the relatively small sample size, for all analyses comparing mean oxy-hemoglobin activation across the three conditions, I adopted a traditional alpha of $p < .05$, but also considered results marginally significant at $p < .1$.

Group ROI Analyses: Focal Right and Left TPJ. Results of the MRI anatomical scan implicated more than one fNIRS channel (channels 2 and 4 on the right, and channels 13 and 16 on the left) as covering the RTPJ and LTPJ ROIs. Given these multiple channels per ROI, the most straightforward and conservative approach is to average data from channels 2 and 4 together, as well as data from channels 13 and 16 together. With a large enough sample, this averaging should wash out differences among the channel pairs, and reveal a robust effect for the right and left TPJ regions of interest.

Omnibus repeated measures analyses of variance (ANOVAs) comparing the mean oxy-hemoglobin activation (averaged across each left and right channel pair) across

Beliefs, Desires, and Physical conditions revealed a significant condition effect in the RTPJ (*F*(2) = 3.67, *p* = .043), and a marginally significant effect in the LTPJ: *F*(2) = 2.65, *p* = .096. Follow-up paired-samples *t*-tests comparing Beliefs versus Desires, Beliefs versus Physical, and Desires versus Physical conditions in the RTPJ and LTPJ were conducted to identify the specific activation patterns across the three conditions, as shown in Figure 4.5a.

Right TPJ. Three paired-samples *t*-tests comparing mean oxy-hemoglobin activation in the RTPJ for the three condition contrasts demonstrated the predicted pattern of activation: mean oxy-hemoglobin (averaged across right TPJ channels 2 and 4) was greater for Beliefs (*M* = .020, *SD* = .022) compared to both Desires (*M* = .011, *SD* = .012) and Physical (*M* = .012, *SD* = .010) conditions (See Figure 4.5a). The difference between Beliefs and Desires conditions was significant (effect size[5] r = .59). Neither the Beliefs versus Physical difference nor the Desires versus Physical difference reached significance or marginal significance. See top panel of Table 4.1 for *t*-test statistics. These results demonstrate that a beliefs-desires distinction exists in the RTPJ, with belief-activation greater than desire-activation, providing evidence that the RTPJ shows specialization for belief-reasoning, beyond desire-reasoning, in children 6- through 10-years-old.

Left TPJ. For the LTPJ, paired-samples *t*-tests of the same three condition contrasts (Beliefs versus Desires, Beliefs versus Physical, and Desires versus Physical conditions) revealed a similar beliefs-desires distinction. Mean oxy-hemoglobin (averaged across left TPJ channels 13 and 16) was greater for Beliefs (*M* = .025, *SD* = .027) compared to Desires (*M* = .014, *SD* = .019), though this difference was only

---

[5] Effect size specific to paired-samples *t*-test calculated as r = square root of (t-squared / [t-squared + degrees of freedom]) as directed by Rosenthal (1991).

marginally significant (and yielded a reduced effect size of r = 0.51 for LTPJ versus 0.59 for RTPJ). Just as in the RTPJ, though the condition difference did not reach significance, activation for the Beliefs condition appeared greater compared to the Physical condition (*M* = .013, *SD* = .015). Also similar to the RTPJ, activation in Desires and Physical conditions did not differ. See Figure 4.5a for pattern of activation across conditions and bottom panel of Table 4.1 for *t*-test statistics. These results suggest that there may also be a beliefs-desires distinction in LTPJ, suggesting a possible diffuse increased recruitment of the TPJ for belief-reasoning, reflecting bilateral specialization for belief-reasoning over and above desire-reasoning. However, any LTPJ effect is less robust compared to the RTPJ results.

Individual ROI Analysis: Focal Right and Left TPJ. An alternative analytic approach to examining neural activation patterns across conditions, and one that is often taken when there is heterogeneity in ROIs across participants, is to use ROIs specific to each individual participant, rather than one general ROI applied to all (e.g., Saxe & Kanwisher, 2003). For the present study, such an individual ROI analysis consists of analyzing data from *either* channel 2 or 4 (on the right), and *either* channel 13 or 16 (on the left) for each participant. The final sample for analysis is thus a composite of data from channels 2/13 and 4/16, with the different singular channels representing each child's individual ROI. To determine which channel represents the best ROI for each child, an initial ROI-inclusion criterion is needed. For the present study, I developed two different criteria, each with their own set of assumptions.

Mental versus Physical ROI criterion. The first ROI criterion was based on an overall general mental-state activation (averaged across Beliefs and Desires conditions) versus physical activation pattern. This criterion yields a more conservative approach to investigating differences between belief- and desire-activations, as it directs where to look for such beliefs-desires distinctions based on a general ‘mental versus non-mental’ criterion. To create individual ROIs for RTPJ, I averaged activation across Beliefs and Desires conditions in channel 2, and did the same in channel 4. I then compared channel 2 ‘mental-state’ activation versus channel 2 physical activation, and channel 4 ‘mental-state’ activation versus channel 4 physical activation. Across each participant, the channel (i.e., 2 or 4) that demonstrated the greatest ‘Mental > Physical’ difference was selected to represent that child’s RTPJ ROI. If a child did not show a Mental > Physical pattern across either channel, they were excluded from the individual ROI analysis. This process was repeated for channels 13 and 16 to define individual LTPJ ROIs. For RTPJ, 8/12 children were included in analyses (5 children with a ch2 ROI and 3 with ch4 ROI). For LTPJ, 9/12 children were included in analyses (3 children with a ch13 ROI and 6 with ch16 ROI).

In both RTPJ and LTPJ ROIs, the predicted pattern of Beliefs activation greater than both Desires and Physical activation was demonstrated. In line with findings from the group ROI analyses, in RTPJ, mean oxy-hemoglobin activation for Beliefs (M = .026, SD = .025) was significantly greater than Desires activation (M = .012, SD = .018), and in LTPJ, the Beliefs-Desires distinction ($M_{Beliefs}$ = .032, SD = .023; $M_{Desires}$ = .019, SD = .013) was again only marginally significant (see Table 4.1, Figure 4.5b). More clearly than in the group ROI analyses, these analyses revealed that Beliefs activation was significantly greater than *Physical* activation in both RTPJ ($M_{Beliefs}$ = .026, SD = .025; $M_{Physical}$ = .003, SD = .020) and LTPJ ($M_{Beliefs}$ = .032, SD = .023; $M_{Physical}$ = .011, SD =

.014,) (see Table 4.1). Moreover, Desires activation (*M* = .019, *SD* = .013) was also significantly greater than Physical activation (*M* = .011, *SD* = .014) in the LTPJ, and was marginally significantly greater than Physical activation in the RTPJ ($M_{Desires}$ = .012, *SD* = .018; $M_{Physical}$ = .003, *SD* = .020) (see Table 4.1, Figure 4.5b).

These results replicate the finding (already demonstrated in the group ROI analysis) that the TPJ is recruited for belief-reasoning, over and above recruitment for desire-reasoning, with the effect demonstrated most robustly in the RTPJ. The confirmation that belief activation is also significantly greater than activation for the Physical condition control is in line with existing findings that children's TPJ is recruited for mental-sate reasoning versus reasoning about non-mental information (e.g., Sommer et al., 2010), and gives further confidence that TPJ demonstrates true specialization for processing beliefs, beyond non-mental processing more generally, and (perhaps most strongly with respect to the RTPJ) beyond processing of other types of mental information such as desires.

Beliefs versus Physical ROI criterion. The second approach to defining individual ROIs was to use inclusion criterion based on a 'Beliefs > Physical' activation pattern, and then focus analyses on investigating specifically the Beliefs versus Desires contrast. This criterion is slightly less conservative than the above individual ROI approach given that it selects a sample with relatively high Belief activations, but it also targets the investigation of neural activations for Beliefs versus Desires more directly. That is, the inclusion criterion focuses examination on regions in which Beliefs activation is already demonstrated strongly (compared to Physical activation), and the question is whether these regions also exhibit strong activation for desire-reasoning, or whether they are recruited for belief-reasoning over and above recruitment for desires.

Similar to the above individual ROI analysis, for the RTPJ, across each participant, the channel (i.e., 2 or 4) that demonstrated the greatest 'Belief > Physical' difference was selected to represent that child's RTPJ ROI. If a child did not show a Belief > Physical pattern across either channel, they were excluded from the individual ROI analysis. This process was repeated for channels 13 and 16 to define individual LTPJ ROIs. For RTPJ, 8/12 children were included in analyses (4 children with a ch2 ROI and 4 with ch4 ROI). For LTPJ, 7/12 children were included in analyses (2 children with a ch13 ROI and 5 with a ch16 ROI).

Results from this individual ROI approach again yielded the predicted pattern of Beliefs activation greater than both Desires and Physical activation in both left and right TPJ. As with the previous two analyses, Beliefs activation (M = .031, SD = .026) was significantly greater than Desires activation (M = .013, SD = .019) in the RTPJ. In the LTPJ, this time the Beliefs-Desires distinction was also significant at $p < .05$ ($M_{Beliefs}$ = .040, SD = .028; $M_{Desires}$ = .016, SD = .012). As expected given that the inclusion criterion maximized Beliefs versus Physical differences, Beliefs activation was also significantly greater than Physical activation in both RTPJ ($M_{Beliefs}$ = .031, SD = .026; $M_{Physical}$ = .007, SD = .023) and LTPJ ($M_{Beliefs}$ = .040, SD = .028; $M_{Physical}$ = .014, SD = .017). Activation in Desires versus Physical conditions did not statistically differ in either ROI. See Table 4.1 and Figure 4.5c. Results of this more targeted analysis again demonstrate significantly greater activation for Beliefs over activation for Desires *and* Physical conditions in the RTPJ, but also demonstrate significance for this same pattern in the LTPJ, suggesting both left and right TPJ may show some specialization for belief-reasoning beyond specialization for reasoning about other types of mental states (such as desires) in our sample of 6- through 10-year-olds.

Contrast AFC ROI. Mean oxy-hemoglobin activation in center channel 8 over the AFC was examined across the three conditions to provide a contrast for the activation patterns in the focal RTPJ and LTPJ ROIs. I examined this region (which is not typically implicated in the theory-of-mind neural network) for activation patterns that differed from those demonstrated in the left and right TPJ. Such differential patterns would give confidence to the focal ROI findings by providing evidence that those focal patterns are specific to their focal regions of interest, and are not merely capturing global activation due to systemic blood flow changes in the scalp and other tissues that lie above the cortex (Boas et al., 2004). Just as for the focal ROIs, I conducted three separate paired-samples *t*-tests comparing activation in Beliefs versus Desires, Beliefs versus Physical, and Desires versus Physical conditions. Results revealed a vastly different pattern of activation: Mean oxy-hemoglobin in the Physical condition ($M = .012$, $SD = .018$) was greater compared to both Beliefs ($M = .002$, $SD = .013$) and Desires ($M = .004$, $SD = .013$) conditions: $t(10) = -2.10$, $p = .063$, effect size r = .55; $t(10) = -2.08$, $p = .064$, effect size r = .55, respectively. Beliefs and Desires activation did not differ from each other: $t(10) = -.57$, $p = .581$. See Figure 4.6 for graphed activation across conditions. This pattern of activation clearly contrasts with the patterns identified in the focal left and right TPJ.

Discussion

The present study investigated the neural correlates of belief- and desire-reasoning, in 6- through 10-year-old typically developing children, using fNIRS methods. The goal was to capitalize on the increased localization capabilities of fNIRS to examine belief and desire neural activations in specifically the left and right TPJ. Results revealed clear and distinct patterns of neural activation across Beliefs, Desires, and Physical (control) conditions. Given that all three conditions had the same perceptual and linguistic structure including similar two-part comparisons, following standard neuroimaging subtraction methodology then, any differences in neural activation across conditions can be attributed to differences in reasoning about the *content* of each condition (i.e., belief-reasoning, desire-reasoning, or reasoning about physical locations) beyond the memory and processing demands common in all conditions.

Thus, in both left and right TPJ, as predicted, neural activation for belief-reasoning was distinctly greater than neural activation for both physical-reasoning *and,* most focally, desire-reasoning. The beliefs-physical distinction was visible across all analyses, and reached significance in two out of three analytic approaches, in both left and right TPJ. The beliefs-desire distinction reached significance across all analyses for the right TPJ, demonstrating a clear, robust belief-specialization on the right. Though visible across all analyses for left TPJ, the beliefs-desires distinction reached significance in only the most targeted analysis, and marginal significance in the other two.

Activation patterns in the contrast region of the anterior frontal cortex give confidence to the focal TPJ results. Activation over that frontal surface region demonstrated a reverse pattern of activation, with physical activation greater than both belief activation and desire activation. Such differential activation patterns in this contrast region indicate that the focal patterns in the left and right TPJ represent true neural activity in the temporoparietal cortex, and are not products of systemic noise, or a global activation pattern occurring over the whole head (Boas et al., 2004). Moreover, given the AFC is not typically implicated in the theory-of-mind neural network, the exhibited reduced activation to the mental-state conditions is in line with our original hypotheses. We had no specific hypotheses about AFC activation for the physical condition. In general, little is known about neural activation associated with the kind of 'diverse-physical' reasoning assessed by our task, and given research suggesting that fNIRS optodes positioned over frontal cortex are less sensitive to near-infrared light absorption changes (Cooper et al., 2012), a deeper interpretation of the AFC findings is not appropriate for the present study. Nonetheless, the contrasting activation patterns exhibited in the AFC region serve their purpose of providing a contrast pattern for the focal left and right TPJ results, and provide the needed confirmation that the patterns exhibited in those focal regions are cognitively meaningful.

The focal findings from the present study have implications both for clarifying the neural correlates of belief- and desire-reasoning specifically, as well as for shedding light on theory-of-mind development generally. I outline these implications next.

Clarifying Neural Correlates of Belief- and Desire-Reasoning

The present study aimed to clarify the results of a parallel ERP study that used our same neuroimaging task with similar-aged children (Bowman et al., 2012). That ERP study demonstrated that in 7- and 8-year-olds, when considering solely the ERP trials for which children performed correctly, neural activation for belief-reasoning differentiated from desire-reasoning in right posterior scalp regions. Our fNIRS findings replicated these ERP findings—providing important validation for the belief-desire reasoning task used in these experiments—and they make two additional, critical contributions. First, the increased spatial resolution and unambiguous localization of the fNIRS data provide clarification that the 'right posterior' belief-desire distinction includes regions of specifically the right TPJ. Second, the fNIRS data demonstrate that not only was right TPJ belief activation different from desire activation activation, it was also clearly *greater* than desire-activation, and moreover, clearly greater than both desire *and* physical activation. These additional clarifications help confirm that the pattern of beliefs, desires, and physical activation in the right TPJ is indeed one of *specialization* for beliefs. That is, our fNIRS results confirm that, by 6- to 10-years-old, specifically the right TPJ shows a more focused and amplified recruitment of neural substrates for belief-reasoning in particular, beyond non-mental processing more generally (evidenced by the belief > physical pattern), and, even beyond processing of other types of mental information such as desires.

One intriguing difference from the Bowman et al. (2012) ERP data is the present study's findings for the left TPJ. Bowman et al. found a belief-desire distinction in only right posterior regions; no difference between belief- and desire-activation was evident on the left. In contrast, fNIRS data revealed a pattern of left TPJ belief-specialization similar to what was found in the right TPJ. The overall effect was less robust, with only the most targeted analysis revealing a significant difference in activation for belief- versus desire-reasoning, yet belief-activation appeared greater than both physical and desire-activation across all three analyses.

Several possible explanations could account for the discrepancy between the ERP and fNIRS left posterior results. The present study sample included slightly younger children (6-year-olds) compared to the 7- and 8-year-olds tested in the ERP study. It is possible that as children age, belief-specialization becomes more focused to the right TPJ, and left posterior specialization for beliefs diminishes, such that the older children in the ERP sample no longer exhibited a left-posterior specialization for beliefs. Findings from existing literature support this possibility. In general, neuro-cognitive activations often become more focused and narrowed with development (Casey et al., 2000). As a most relevant example, Saxe et al. (2009) measured neural activation (via fMRI) as typically developing 6- through 10-year-olds listened to stories describing peoples' mental states (mental condition), peoples' interactions and appearances (social condition), and physical objects and scenes (physical condition). Results showed greater activation in the mental condition versus the social condition in both the left and right TPJ, but as children aged, specifically the right TPJ was found to increase in selectivity for mental-state processing (in comparison to processing both physical and social stories). Thus, the absence of younger children in the ERP study could have resulted in the reduction of a left-side belief-specialization effect to the point where the ERP analyses could not detect it.

Likewise, the particular characteristics of the fNIRS data and methods—the increased spatial resolution, clearer measures of individual variation in neural responses, and generally more targeted analytic approach to left TPJ examinations—could have been necessary to reveal the weaker left posterior specialization, and thus the ERP analyses in Bowman et al. (2012) may not have been able to reveal such a left-side effect, even if it was present. Indeed, in the fNIRS study, only the most targeted analytic approach yielded a significant left TPJ specialization for beliefs. Under this explanation, our fNIRS data offer an important extension of the ERP study results, demonstrating specialization of belief-reasoning beyond desire reasoning in the right TPJ, as well as some evidence for (perhaps a less robust) specialization on the left—a pattern in line with existing investigations of theory-of-mind reasoning that also use hemodynamic data as described above (e.g., Saxe et al., 2009). Our limited sample size did not allow investigation of whether left and right belief-specialization changed as a function of age. Future fNIRS investigations of belief- and desire-reasoning in older children, as well as in children younger than 6 years, and with overall greater sample sizes are important to further explore this intriguing potential developmental effect.

Broader Implications for Theory-of-mind Development

The present study provides clear evidence that by as early as 6-years-old, neural regions in the right TPJ and (to a lesser degree) in the left TPJ exhibit specialized recruitment for belief-reasoning, beyond not only non-mental reasoning, but also beyond *desire*-reasoning. As discussed above, these results replicate and extend results of a parallel investigation of belief- and desire-reasoning in similar-aged children using ERP. In particular, the now robustly demonstrated pattern of specialization in the right TPJ has implications for understanding broader aspects of theory-of-mind development.

In Bowman et al. (2012), the authors suggest that the belief-desires distinction in right posterior regions, that was exhibited only when analyses were concentrated on correct performance trials (and not evident when incorrect trials were included in analyses), suggests a trend in specialization of right posterior regions for specifically belief-reasoning that is linked to increasing accuracy for mental-state and belief-state reasoning. Indeed, evidence from recent investigations of general mental-state reasoning in children provide support for this argument; in brief, fMRI research demonstrates that specifically the right TPJ increases in specialization for processing mental-state information (versus physical information and general social information about interactions and appearances) as children age from 6- to 10-years old (Saxe et al., 2009), and as accuracy for behavioral theory-of-mind reasoning improves over the same time period (Gweon et al., in press). The fNIRS data add the crucial pieces that connect findings from Bowman et al. to these fMRI findings. Specifically results from the present study suggest that the extant fMRI findings for right TPJ specialization could actually be evidencing specialization not only for mental-state reasoning versus non-mental reasoning (e.g., social- and physical-reasoning), but specialization for *belief*-reasoning specifically, beyond specialization for other types of mental-states such as desires.

This pattern of specialization has clear implications for understanding the behavioral findings demonstrating that children reach an explicit understanding of desires before they come to an explicit understanding of beliefs (e.g., Wellman & Liu, 2004), pointing to a neural mechanism underlying this progression. That is, results from the present study (in conjunction with Bowman et al., 2012) represent a straightforward developmental possibility that an understanding of beliefs may build on prior desire-understanding, evidenced by additional substrates in right TPJ regions being recruited for specifically belief reasoning—beyond recruitment for desire-reasoning—as belief-understanding becomes more distinct and accurate. Such a developmental scenario is supported by behavioral evidence that children progress from an explicit understanding of desires to an explicit understanding of beliefs, (e.g., Wellman, 2002; Wellman & Liu, 2004), as well as data from the current study. Indeed, our performance data show that even by middle childhood, children still have greater accuracy on diverse-desires tasks compared to diverse-beliefs tasks (although children in our study are older than ages when children, on average, pass diverse-desires and diverse-beliefs tasks, our tasks presented information quickly and thus were more difficult overall)—a pattern consistent with findings demonstrating that belief-understanding can be less accurate/fluent in older children and adults (e.g., Malle, 2004).

Findings from the present study thus highlight the utility of fNIRS data for identifying neural specialization in targeted regions of interest, as well as the utility of neural data in general for shedding light on behavioral research findings. Of course, future research is needed to directly test the above developmental scenario, especially longitudinal research to determine whether neural specialization might precede increased accuracy or vice versa. An important first step is to collect more data using the present study's tasks and methods. I outline future directions specific to our fNIRS approach below.

Next Steps for fNIRS Investigations of Belief- and Desire-reasoning

As an immediate next step, additional child data should be added to the present study for further analysis. The central advantage to an increased sample concerns direct tests of relations between performance accuracy and neural specialization for belief-reasoning. That is, though the present study was set up to be able to more directly investigate the extent to which increased neural specialization relates to increased accuracy for belief- and mental-state reasoning, such analyses were inappropriate for the current small sample size. When data from additional participants are added to the current study, the clearer measures of individual variation in neural responses (i.e., percent signal change in oxyhemoglobin response) afforded by fNIRS methods should allow correlation analyses between performance accuracy and neural activation across conditions. These correlation analyses will directly investigate the possibility that a growing distinction between belief- and desire-activations in the TPJ is related to an increasing accuracy for mental-state reasoning. Thus, an additional 12 children within the current study's age range should be tested. This doubled sample size should offer reasonable power and variance sufficient for correlation analyses, and should also clarify some of the condition effects that reached only marginal significance (i.e., in the left TPJ in particular). Moreover, given that this second set of data would likely be collected with a different fNIRS acquisition system, equal-sized samples can best test for any between-sample differences that may exist due to differences in experimenters, or acquisition systems. A specific focus on recruitment of 7- and 8-year-olds is recommended in order to most closely parallel the ERP sample in Bowman et al. (2012)

Data from different age groups should also be collected using our beliefs-desires fNIRS task. In particular, data from adults would allow comparison to the parallel adult ERP data in Liu et al. (2009a) and thus could provide additional confirmation for our task and methods, as well as allow developmental comparison between child and adult findings, as was done in Bowman et al. (2012). Data from younger children (e.g., 3- to 5-year-olds) would also be particularly valuable in shedding light on the hypothesis that TPJ specialization for beliefs is associated with belief- and mental-state reasoning accuracy. Because of the child-friendly features of fNIRS (in contrast to fMRI), fNIRS methods could be particularly valuable for research in this preschool age range.

Moreover, children in this preschool age-range vary in performance on the standard behavioral measures of explicit desires and explicit beliefs. Thus, investigations of how neural correlates of belief- and desire-reasoning develop over this time, and how they might relate to performance on these standard behavioral measures, would be especially intriguing. Adult and preschool samples would also provide a sufficient age spread to shed light on the possibility that left TPJ specialization for belief-reasoning diminishes with age.

Summary

In advance of useful future research, the present study already sheds light on the neural correlates of belief- and desire-reasoning in middle childhood. Results reveal that, in 6- through 10-year-old children, the TPJ exhibits specialization for belief-reasoning, beyond both non-mental physical reasoning as well as beyond reasoning about other types of mental states such as desires. These results clarify those of a parallel study that examined belief- and desire-reasoning in similar-aged children using ERP (Bowman et al., 2012). Taken together, findings from the present study and Bowman et al., in conjunction with existing fMRI research (Saxe et al., 2009; Gweon et al., in press), point to a possible neural mechanism underlying the developmental progression from understanding desires to understanding beliefs evidenced by numerous behavioral studies (e.g., Wellman & Liu, 2004). Specifically, an explicit understanding of beliefs may build off prior understanding of desires, evidenced by an increased specialization in the TPJ for reasoning about specifically beliefs (over and above desires) as this reasoning becomes more distinct and accurate. Future research, using fNIRS methods in particular, can now further address this hypothesis, to shed more light on how different types of mental states build to form an expert theory of mind.

Table 4.1

Paired-Samples T-Tests Comparing Children's Mean Oxy-Hemoglobin Response in the Three Condition Contrasts in Left and Right TPJ for the Three Analytic Approaches

Comparison	Analytic Approach		
	Group ROI (Channel-pair Average)	Individual ROI: Mental > Physical	Individual ROI: Beliefs > Physical
		RTPJ	
Belief vs. Desire	*t*(10) = 2.32, *p* = .043 *	*t*(7) = 2.42, *p* = .046 *	*t*(7) = 4.20, *p* = .004 *
Belief vs. Physical	*t*(10) = 1.75, *p* = .110	*t*(7) = 2.78, *p* = .027 *	*t*(7) = 3.14, *p* = .016 *
Desire vs. Physical	*t*(10) = -.46, *p* = .653	*t*(7) = 2.16, *p* = .068 ✝	*t*(7) = 1.32, *p* = .224
		LTPJ	
Belief vs. Desire	*t*(10) = 1.87, *p* = .091 ✝	*t*(8) = 1.89, *p* = .095 ✝	*t*(6) = 3.72, *p* = .010 *
Belief vs. Physical	*t*(10) = 1.56, *p* = .149	*t*(8) = 2.78, *p* = .024 *	*t*(6) = 4.22, *p* = .006 *
Desire vs. Physical	*t*(10) = -.33, *p* = .784	*t*(8) = 3.88, *p* = .005 *	*t*(6) = 1.26, *p* = .254

Notes. * indicates significance at $\alpha = .05$. ✝ indicates significance at $\alpha = .1$.

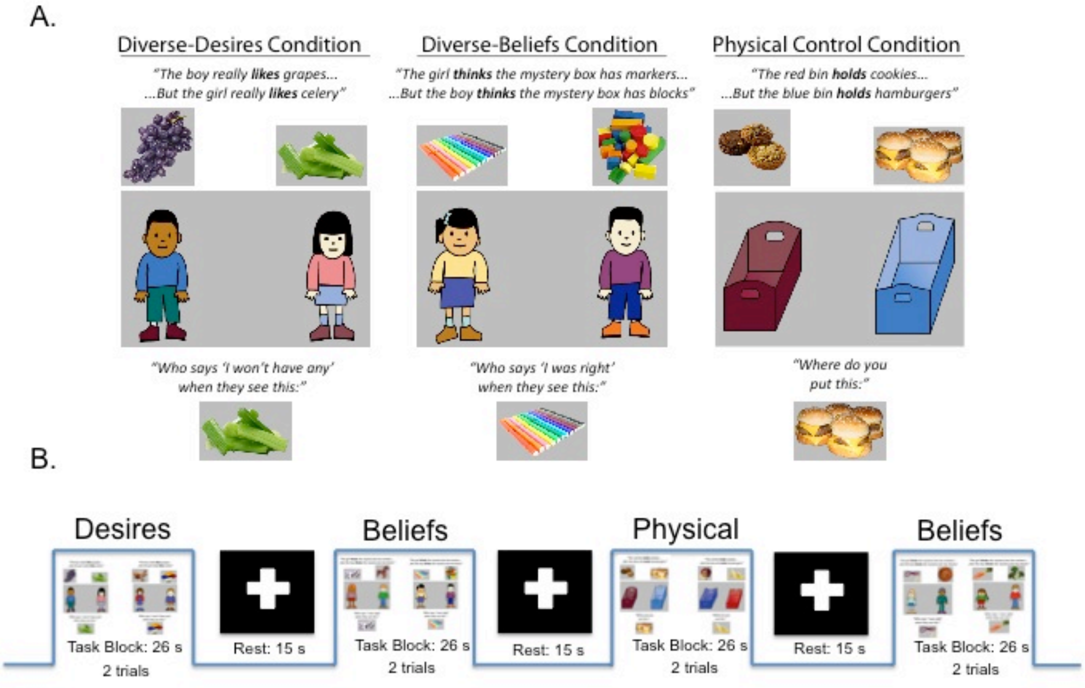


Figure 4.1. (Adapted from Bowman et al., 2012). (A) Examples of single trials for Diverse-Desires condition (left), Diverse-Beliefs condition (middle), and Diverse-Physical condition (right) with examples of information phase (top) and target questions (bottom) as well as sample graphics for both food and toy trial types. (B) Schematic of experimental and rest blocks as presented within a run.

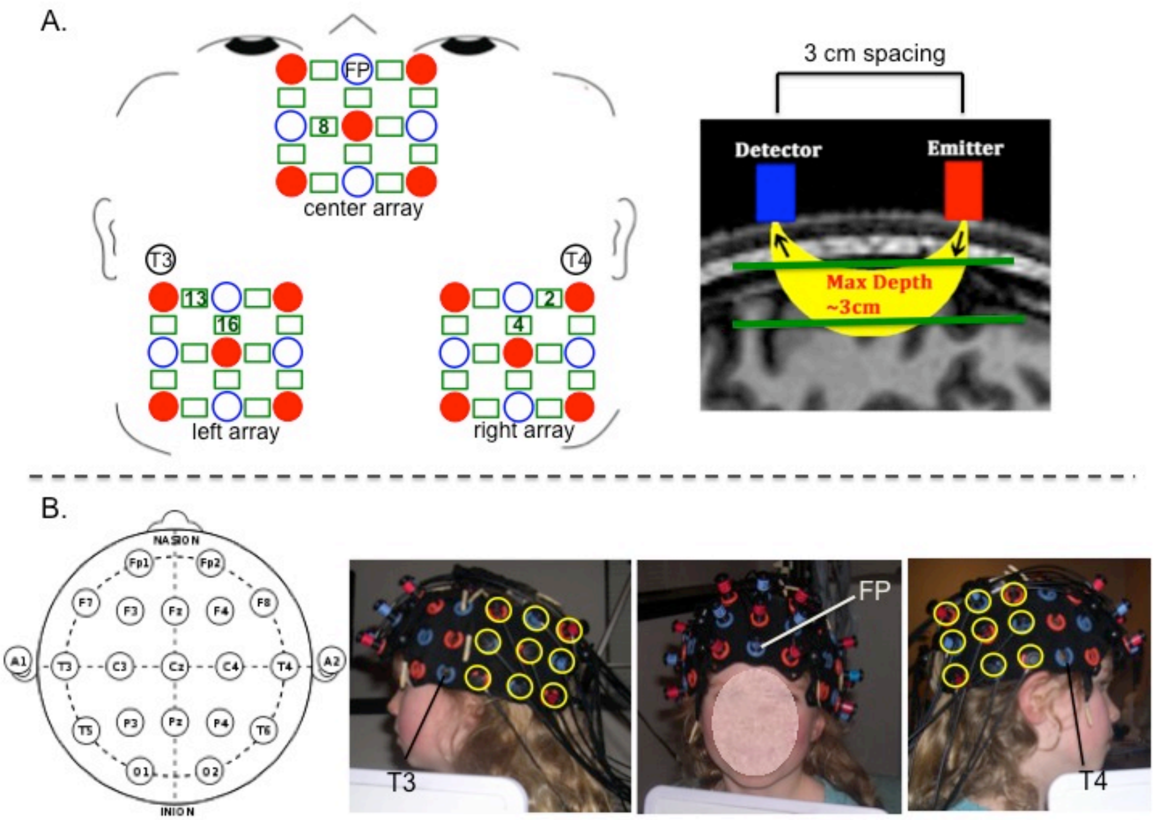


Figure 4.2. (A-left) Schematic of center, left, and right optode arrays showing emitters as red solid circles, detectors as blue open circles, and channels as green rectangles. 10-20 landmarks corresponding to specific optode placements are labeled. Green numbers within rectangles mark the positions of the specific channels used for analyses. (A-right) Depiction of spacing between detector and emitter optodes and channel of near-infrared light penetrating 3cm through cortex. (B) Photographs of optode placement on child participant guided by 10-20 landmarks.

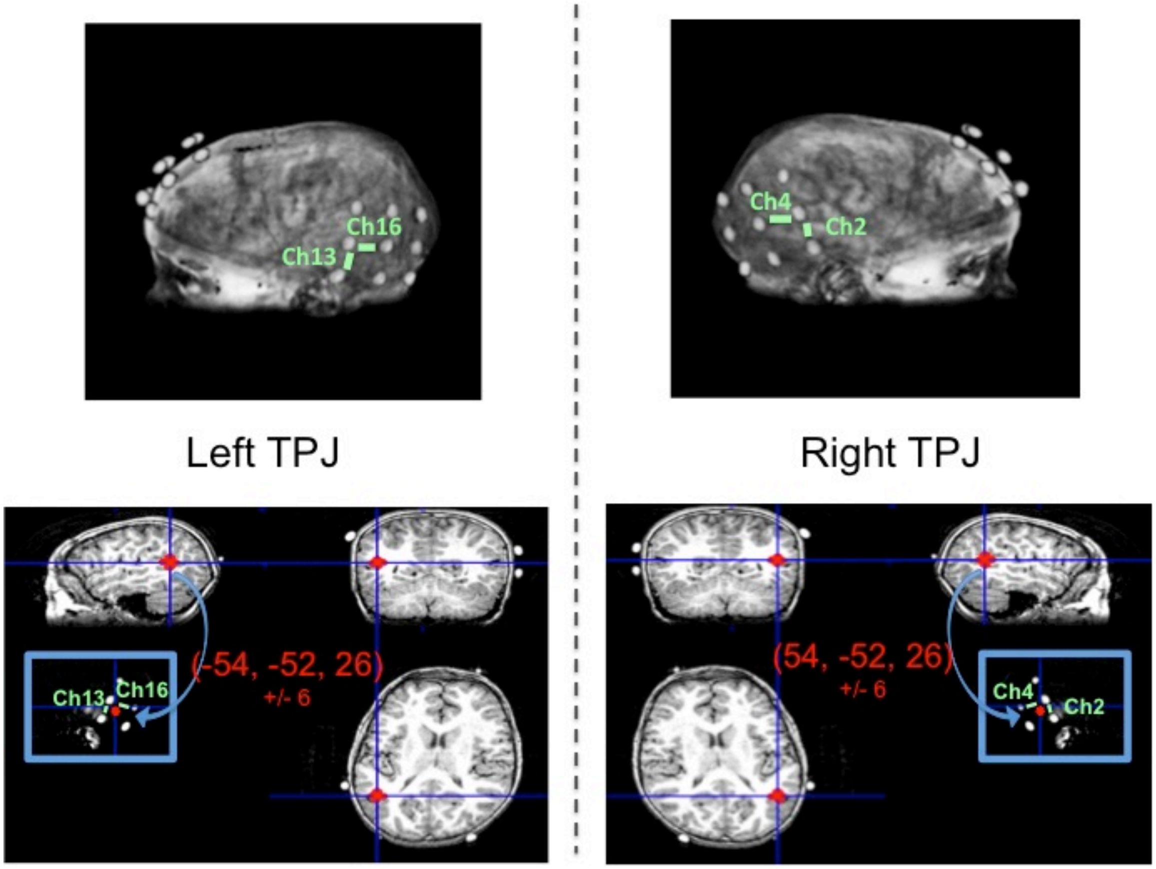


Figure 4.3. MRI anatomical scan of representative child participant with vitamin E probes secured to head mimicking optode positions (top left and right). Focal channels of interest are labeled for the left TPJ (top left) and right TPJ (top right). Bottom panel depicts the TPJ regions of interest (red circles) that include the range of coordinates for left TPJ (bottom left) and right TPJ (bottom right) as identified in a separate theory-of-mind fMRI study with children of similar ages to the present study.

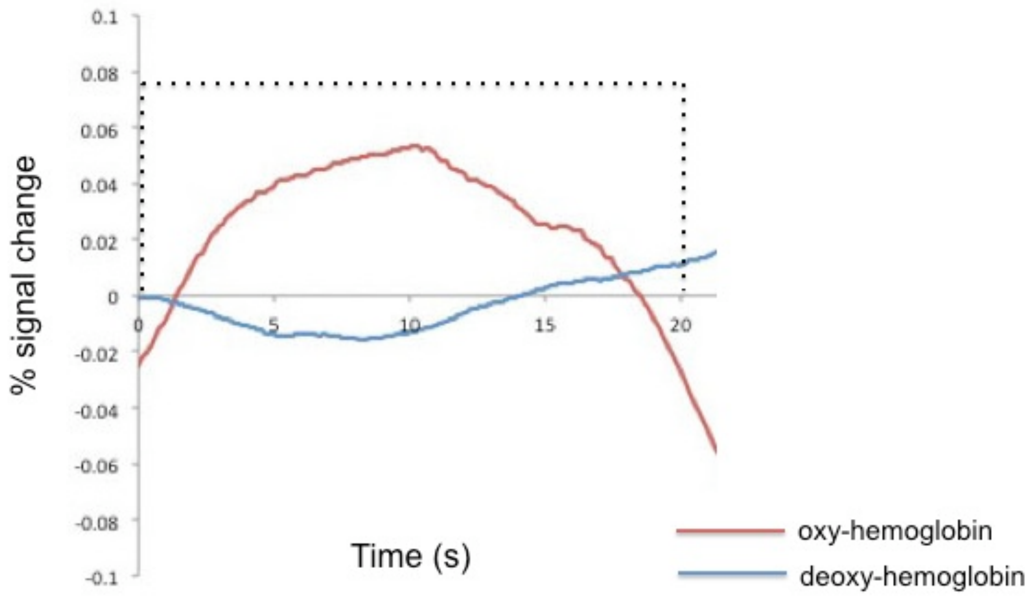


Figure 4.4. Example of appropriate oxy- (red) and deoxy- (blue) hemoglobin response curves averaged across experimental blocks for a single participant. Black dotted line maps out the time window in which data were analyzed for the focal left and right arrays. The curves exhibit the canonical slow surge in oxy-hemoglobin with simultaneous decrease in deoxy-hemoglobin over the course of the experimental block before returning to baseline.

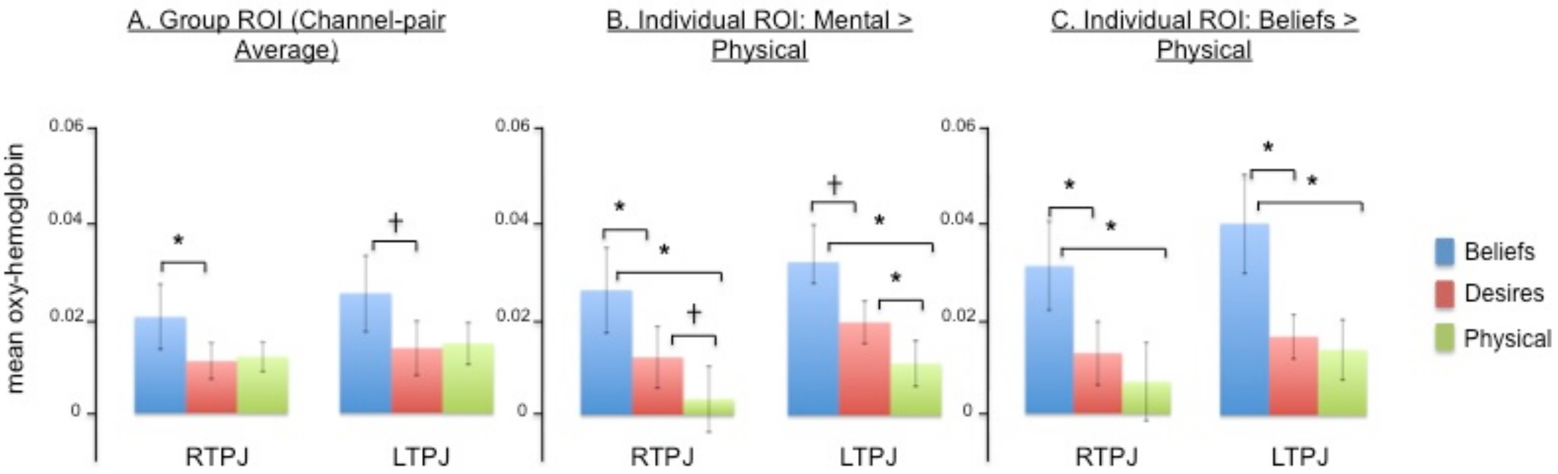


Figure 4.5. Percent mean oxy signal change (from rest) for beliefs conditions (blue), desires conditions (red), and physical conditions (green) in LTPJ and RTPJ for the group ROI analyses (A), the individual ROI analyses with 'Mental > Physical' inclusion criteria (B), and for the individual ROI analyses with 'Beliefs > Physical' inclusion criteria (C). Results indicate specialization for belief-reasoning (over desire-reasoning) in the RTPJ as evidenced by significantly greater oxy signal for beliefs versus desires conditions across all three analyses. Results also suggest some evidence for belief-specialization in the LTPJ, though effects are less robust. * $p < .05$, † $p < .1$.

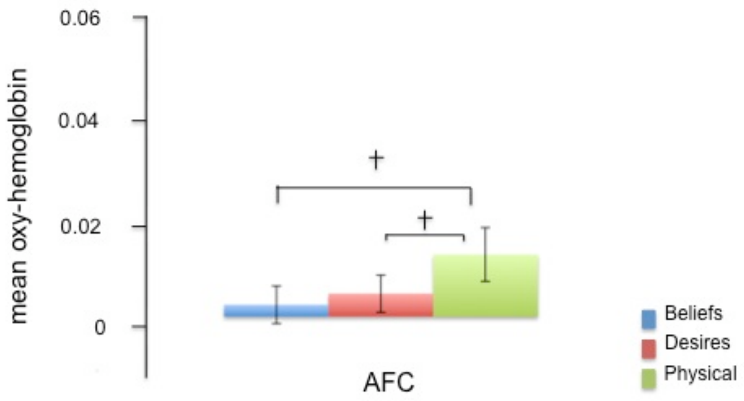


Figure 4.6. Percent mean oxy signal change (from rest) for beliefs conditions (blue), desires conditions (red), and physical conditions (green) in channel 8 over the AFC control region. Results indicate a pattern of activation that contrasts with results for the focal left and right TPJ. † $p < .1$.

CHAPTER V

Study 3. Neural Correlates of Belief- and Desire-Reasoning in Late Childhood: An Event-Related Potential Study

Introduction

Having a "Theory of Mind"—the understanding that people's actions are guided by internal mental states such as beliefs, desires, and intentions—is a cornerstone of social-cognitive development (Harris 2006; Wellman, 2002). Theory of mind goes beyond social perception; it requires conceptualization of and reasoning about people's mental states in order to accurately predict and explain behavior.

Important achievements in development of theory of mind are manifest during the preschool years (e.g., Wellman & Liu, 2004). An often-studied, clear example of these achievements is the development of false-belief understanding; for example, understanding that a boy could believe his dog was in the shed when in reality it was at the playground. Numerous studies show that children transition from consistently failing to consistently passing standard false-belief tasks in the preschool and early school years (e.g., Wellman, Cross, & Watson, 2001), demonstrating development of an explicit understanding of beliefs and of the mind more generally. However, focus on beliefs and false beliefs alone is limited. Theory of mind—often termed a belief-desire or belief-desire-emotion naïve psychology—involves understanding multiple causally interconnected mental concepts, and developmentally, children's theory of mind proceeds in a progression of mental-state understandings.

A crucial, well-documented progression is that children consistently develop an explicit understanding of desires before developing an explicit understanding of beliefs—they can understand that the boy *wants* to find his dog before understanding that he *thinks* it is in the shed (e.g., Bartsch & Wellman, 1995; Gopnik & Slaughter, 1991; Wellman & Liu, 2004). The purpose of the present study is to examine, using event-related brain potential (ERP) measures, the neural mechanisms underlying understanding of both desires and beliefs in 10- and 11-year-old children—thereby providing insights into children's developing understanding of the mind.

The extant findings of understanding desires before beliefs hold across tasks matched on procedural methodology, linguistic structure, and materials. For example, one can directly compare diverse-desires versus diverse-beliefs tasks (used by Wellman & Liu, 2004; and also Wellman & Woolley, 1990). For the diverse-desires task, children are told about a character who *likes* a particular snack opposite to what the child prefers (e.g., likes carrots not cookies). Children are asked to predict which food the character will choose for snack (the carrot or the cookie). For the matching diverse-beliefs task, children are told about a character who *thinks* his cat is hiding in a location opposite to what the child thinks (e.g., thinks the cat is in the garage but child thinks the cat is in the tree). Children are asked to predict where the character will look for his cat (in the garage or in the tree). The demands and format for the two tasks are virtually identical except that children must predict behavior based on different mental concepts (desires versus beliefs). Children consistently pass diverse-desires tasks at an earlier age (e.g., around age 2 years) than diverse-beliefs tasks (e.g., age 3 and 4 years). Recent looking-time studies on infants' false-belief understanding (e.g., Onishi & Baillargeon, 2005) have added fuel to the debate about the age at which children possess the capacity to process others' beliefs, at least implicitly. Nevertheless, studies of explicit performance strongly support a consistent developmental progression of explicit understanding of desires before explicit understanding of beliefs. Indeed, the desire-belief progression is present across different cultures (e.g., Wellman, Fang, Liu, Zhu & Liu, 2006), and it even occurs in populations in which social development is impaired (e.g., in late-signing deaf children and high functioning individuals with autism; Peterson et al., 2005).

Yet, despite the wealth of existing behavioral research, the nature of this cognitive-developmental progression remains unclear. This lack of clarity is in part due to the fact that children's performance reaches ceiling on the standard behavioral theory-of-mind tasks at roughly age 6 or 7 years (see e.g., Wellman et al., 2001; Wellman & Liu, 2004). Belief- and desire-reasoning surely develop beyond these ages, but compared to behavioral research in early childhood, theory-of-mind research in older children and adults is thin; in general there are fewer studies that examine theory-of-mind reasoning in these older populations, and the behavioral measures used vary greatly, making developmental comparisons difficult (e.g., Miller, 2012).

However, some research does show that aspects of belief-reasoning continue to be less fluent and accurate in older children and adults (e.g., Malle, 2004). For example, adults make errors when interpreting another person's references and descriptions of objects when their own beliefs and knowledge about the object conflict with the beliefs of the other person (Keysar et al., 2003). And adults and children 6- to 11-years-old show lower accuracy and longer response times for reasoning about false-beliefs versus true-beliefs, suggesting greater difficulty with the former versus latter type of reasoning (Apperly, Warren, Andrew, Grant & Todd, 2011). Very little is known about desire-reasoning in older populations. Some studies exist that measure belief- and desire-reasoning together, requiring participants to make judgments about a character's belief given the character's preference or desire (e.g., Apperly et al. 2011; Friedman & Leslie, 2004). However, behavioral measures in older children and adults that directly contrast belief-reasoning with desire-reasoning are virtually nonexistent. Thus, much is unknown about a) the developmental trajectories of these two types of mental-state reasoning and how they are related beyond early childhood, and b) the mechanisms and processes that underlie the progression of understanding desires to understanding beliefs.

Neuroscientific research can shed light on the neural processes underlying cognitive development. Indeed, much is now known about the neural correlates of theory-of-mind reasoning in adults. Functional neuroimaging and electrophysiological studies with adults converge on findings that theory-of-mind reasoning recruits a network of neural regions most consistently including the medial prefrontal cortex (MPFC) and temporoparietal junction (TPJ) (see Carrington & Bailey, 2009; Apperly, 2011, for reviews). Some of this research suggests that the TPJ, and perhaps the *right* TPJ in particular might be especially recruited to process information about complex mental-states (e.g., beliefs). Saxe and Wexler (2005) demonstrated that adults selectively recruited the right TPJ for processing mental states but not for processing other socially relevant facts about a person (i.e., marital status, family relations, cultural background), and that none of the other regions typically recruited in theory-of-mind reasoning (i.e., medial prefrontal cortex) showed such a specified role. Similarly, fMRI data shows that for adults, TPJ is recruited for belief-processing over processing of physical or general social characteristics (i.e. a person's appearance) (Saxe & Powell, 2006), and that the belief-processing that occurs in the right TPJ in particular is specific to belief-reasoning, and independent of domain-general processing of executive functioning demands (Saxe, Schulz, & Jiang, 2006).

Recent research is beginning to examine neural correlates of theory-of-mind reasoning in children as well. This research demonstrates that both the MPFC and TPJ are also important for theory-of-mind reasoning in childhood and early adolescence. For example, the TPJ and MPFC are both recruited when children 11- to 14-years-old think about their own and others' beliefs (Pfeifer et a., 2009), and when children 8- to 11-years-old reason about cartoons depicting characters' true and false beliefs (Sommer et al. 2010; Kobayashi, Glover, & Temple, 2007). Recent EEG research demonstrates that even in children as young as 4-years-old, the MPFC and right TPJ are important for theory-of-mind reasoning: Functional maturation of each of these regions positively correlated with children's behavioral theory-of-mind performance, independent of any correlations with age or domain general executive functioning and language skills (Sabbagh, Bowman, Evraire, & Ito, 2009).

Similar to adults, some research suggests that the right TPJ may play a particularly important role in theory-of-mind development. Saxe et al. (2009) measured neural activation (via fMRI) as typically developing 6- through 10-year-olds listened to stories describing peoples' mental states (mental condition), peoples' interactions and appearances (social condition), and physical objects and scenes (physical condition).

Results showed greater activation in the mental condition versus the social condition in the MPFC, and in both the left and right TPJ, but as children aged, specifically the right TPJ was found to increase in selectivity for mental-state processing (in comparison to processing both physical and social stories). Gweon et al. (in press) replicated these results with children 5- to 11-years-old using similar mental, social, and physical stories, but further demonstrated that the right TPJ increased in mental-state selectivity as children's behavioral theory-of-mind reasoning improved.

Thus, neuroscientific research in both adults and children consistently implicates the MPFC and TPJ as important for theory-of-mind reasoning, with the right TPJ potentially playing a particularly important role in theory-of-mind development; as children age and become more accurate at theory-of-mind reasoning, the right TPJ exhibits an increased specialization for theory-of-mind reasoning specifically (e.g., Saxe et al., 2009; Gweon et al., in press). However, a key element is missing in both adult and child research that would allow for a more complete window on the neurological mechanisms underlying theory-of-mind development. Specifically each of the above studies have focused on either belief-reasoning specifically, or mental-state reasoning generically—a most common practice in adult and child neuroscientific examinations of theory-of-mind in general. To reiterate, theory of mind is often termed a belief-desire naïve psychology to signal that it involves understanding multiple causally interconnected mental concepts (in particular both belief-states—that represent the world, and desire-states—that motivate particular actions within the represented world, to get what one wants). Moreover, the most prominent and substantiated developmental progression is one from robust early desire-understanding to later belief-understanding.

Therefore, much can be learned about theory-of-mind development from neuroscientific investigations of multiple mental-state understandings; critically, examinations of *desires* as well as beliefs.

One recent study has directly examined the neural correlates of belief- *and* desire-reasoning in children. Bowman, Liu, Meltzoff, and Wellman (2012) recorded ERPs when 7- and 8-year-olds performed diverse-desires and diverse-beliefs tasks, as well as physical reasoning tasks as a control condition. The tasks, methods, and EEG acquisition system were identical to those used in the Liu et al. (2009a) ERP study with adults, to enable direct comparisons between child and adult ERP data. Both adults and children showed frontal neural activations for belief- and desire-reasoning that equally differentiated from activations for the physical control, but did not differentiate from each other—suggesting a common system for reasoning about mental states more generally (including both beliefs and desires) in mid-frontal regions in both age groups. In right posterior scalp regions, adults also showed neural activation that was specific to belief-reasoning, distinct from both physical *and* desire-reasoning. This pattern was only partly observed in children. When all child ERP trials were included in analyses (as was done for the adult data), no right posterior belief-desire distinction was evident. Only when analyses were concentrated on the trials in which children judged correctly did right posterior activations for belief-reasoning emerge as distinctly greater than activations for desires.

The authors reason that these data suggest a growing specialization of the right posterior brain regions for specifically belief-reasoning, pointing to a mechanism whereby developmental increases in accuracy for inferring complex mental states contribute to the development of neural specialization that supports social-cognitive understanding. Moreover, they suggest a possible explanation for the progression of understanding desires to understanding beliefs seen in children: children may need to recruit additional neural processes (within right posterior regions) for reasoning about beliefs beyond a common neural system (within mid-frontal regions) for reasoning about mental states more generally.

These hypotheses rest on comparisons between child and adult ERP activation patterns for belief- and desire-reasoning, which on a broad level are comparable. There were, however, some intriguing differences that existed between the 7- and 8-year-olds' ERP activation patterns (Bowman et al., 2012) and the ERP patterns observed in adults (Liu et al., 2009a). Specifically, there were four key differences between adult and child data. The first three differences concern the mid-frontal findings: 1) Though both adults and children showed equal differentiation of belief- and desire-activation patterns from the activation patterns for the physical control, children exhibited mental activations that were more *negative* compared to activation for the physical condition, whereas adults exhibited mental activation that was more *positive* compared to the physical condition. 2) Though both adults and children demonstrated this effect in mid-frontal scalp regions, in children the effect was more diffuse and concentrated more heavily over the *right* mid-frontal scalp versus the scalp effect in adults, which concentrated solely over the center-frontal regions. And 3) for adults, the mental-versus-physical effect occurred only in the later epochs beyond 800 ms post-stimulus, whereas children demonstrated this effect in later *and* earlier epochs, beginning as early as 200 ms and sustained beyond 800 ms post stimulus.

The final difference concerned the right posterior beliefs-versus desire distinction: Both children and adults demonstrated significantly more positive activation for beliefs compared to desires in the 600-800 ms epoch, however children also showed this same effect in the later epochs of 800-1400 ms post stimulus. See Figure 5.1 for visual summary of similarities and differences between adult and child ERP data.

The Present Study

A deeper examination of child-adult differences in ERP activations in its own right, and in regard to the features just outlined, is needed. Such an examination could shed more light on the *development* of neural correlates for belief- and desire-reasoning, and help clarify the processes underlying how different mental-state understandings build to form an adult theory-of-mind. The present study seeks this further clarification by using the same tasks and stimuli used by Liu et al. (2009a) and Bowman et al. (2012) (diverse-desires and diverse-beliefs tasks with physical reasoning tasks as control) to collect ERP data from typically developing 10- and 11-year old children.

Children 10- and 11-years-old are two to four years older than the 7- and 8-year-old sample, but are still young enough to contrast developmentally with the adult data. Crucially, these older children should exhibit a higher accuracy for belief- and desire-reasoning compared to the younger children from Bowman et al. (2012). Thus, from a purely behavioral perspective, behavioral performance data from the present study can be compared to performance data in the 7- and 8-year-old sample to examine how accuracy for belief- and desire-reasoning improves as children age beyond early childhood—beginning the process of filling the distinct void of information on belief- versus desire-reasoning in these older age ranges. Moreover, critically, a comparison of ERP activation patterns between the present study's 10- and 11-year-olds and the younger 7- and 8-year-olds allows for an examination of how belief and desire neural correlates might change as accuracy for belief- and desire-reasoning improves.

With respect to the ERP data, several informative patterns of results might appear. It is possible that ERP patterns for the present study's 10- and 11-year-old children will remain the same as those identified in the younger 7- and 8-year-old children (Bowman et al., 2012), thereby suggesting continuity in neural correlates of belief- and desire-reasoning over middle to late childhood. It is also possible that ERP patterns in the present study will directly match the patterns identified in adults (in Liu et al., 2009a), suggesting rapid development of neural systems supporting belief- and desire-reasoning in the two to four years beyond middle childhood. Finally, it is possible that ERP activations for the present study's 10- and 11-year-olds may retain some similarities to the younger sample, but also exhibit more similarities to the adult data, thereby demonstrating a transitional state, and capturing development in the neural systems supporting belief- and desire-reasoning as children age into late childhood. How these similarities and differences turn out in terms of the four discrepancies between the adult and 7- and 8-year-old samples outlined above (i.e., discrepancies in location, polarity, and timing of effects) could help shed light on the processes by which neural correlates for belief- and desire-reasoning change over middle to late childhood, and how these neural correlates support the development of an expert and complete theory-of-mind.

Methods

Participants

Fifty-six 10- to 11-year-old typically developing children (34 males) participated in the study. All participants were right handed, with normal or corrected-to-normal vision. One participant was excluded due to abnormally high temperatures in the testing room causing prominent sweat artifact in the EEG data. Thirty-five additional children did not provide at least 15 usable, artifact-free trials of electrophysiological data for each of the three conditions, yielding a final sample of 20 participants (age range = 120-142 months; *M* = 129.45 months, *SD* = 7.32; 12 males) used in our full-trial analyses (as described below in the results section). Upon closer inspection of exclusions, it was found that 82% of the total exclusions occurred during tests by novice experimenters. A higher exclusion rate from novice experimenters is common for electrophysiological research with children (e.g., Sabbagh et al., 2009). The more experienced experimenters were able to collect usable data from 64% of participants tested, representing an inclusion rate in line with existing electrophysiological studies that used the present ERP task (Bowman et al., 2012; Liu et al., 2009a), and in line with existing pediatric electrophysiological research (Sabbagh et al., 2009).

Stimuli and Procedure

To parallel Bowman et al. (2012) as closely as possible, we used the same tasks, stimuli, EEG recording system, and conditions: multi-trial diverse-desires, diverse-beliefs, and diverse-physical judgment tasks created to collect ERPs from participants. Descriptions of stimuli and procedures are taken directly from Bowman et al.

For the ERP task, the structure of all 144 trials (48 in each of three conditions) was the same. In each trial, participants first received information about two characters with different desires for food/toys, two characters with different beliefs about food/toys, or two locations to put food/toys away. Participants read information in text that was presented on screen (e.g., “the boy likes grapes, but the girl likes celery”) with accompanying pictures. An experimenter sat beside the participant and after the first block asked participants if they would like help reading the text. Two participants asked for help, and so for the remaining trials for those participants, the experimenter read all text aloud as it was presented on screen. All other participants read the text on their own. On a random third of trials, the initial information was followed by a memory check to ensure participants paid attention to each trial (e.g., “who likes grapes?”). If participants answered the memory question incorrectly, the information phase was repeated. After the information phase of each trial (7300 ms in duration), participants read the target question (details provided below for each condition) and were presented with a picture of one of the two food/toys (e.g., celery) for 2000 ms. This pictorial presentation of a single food/toy was the target visual event to which ERP data were time-locked. Participants then answered the target question (via button press). See Figure 5.2 for schematic summary of the tasks.

Desire-reasoning. For the Desires condition, in each trial, participants read about a boy who likes a particular food/toy (e.g., grapes/markers), a girl who likes a different food/toy (e.g., celery/blocks), and a closed box said to contain either a snack or a toy (with boy or girl first being counter-balanced). Participants were then presented with one of four target questions about what would happen when the box was opened: “Who says

'I'll have some' when they see this?" / "Who says 'I won't have any' when they see this?" (when the story was about food), or "Who says 'I'll play with it' when they see this?" / "Who says 'I won't play with it' when they see this?" (when the story was about toys). The question presented (either positive or negative wording) was randomized in each trial. After the target question, participants were immediately presented with a picture of one of the two foods/toys (e.g., celery/blocks). Note that in this and other conditions, participants were not able to answer the target question until presented with a picture of the food/toy that was in the closed box. After 2000 ms of seeing the revealed food/toy, participants answered by choosing one of the two characters.

Belief-reasoning. The Beliefs condition followed the same presentation format (including counterbalancing and randomization) as just described, except the content of the information and target questions were about beliefs. Participants read about a mystery box containing food/toys for a guessing game, and were introduced to a girl who thinks the box contains a particular food/toy and a boy who thinks the box contains a different food/toy. The target questions for this condition were: "Who says 'I was right' when they see this?" or "Who says 'I was wrong' when they see this?" After the target question, participants were presented with a picture of one of the two foods/toys, and then answered by choosing one of the two characters.

Physical-reasoning. The Physical condition provided a non-mental control condition (again following the same presentation format as the other two mental-state conditions). Participants read about a closed box containing food/toys to put away, and were informed that the red bin should receive a particular food/toy while the blue bin should receive a different food/toy. Target questions for this condition were: "Where do you put this?" or "Where do you not put this?" After the target question, participants were presented with a picture of one of the two foods/toys that was in the closed box, and answered by choosing one of the two bins.

Commonalities across conditions. Because trials for all three conditions were constructed to have the same perceptual and linguistic structure, any differences between conditions would point to the *mental-state processing* beyond these perceptual and task similarities.

Trials for each condition and type (e.g., desires condition about toys) were presented in blocks of six. Blocks semi-randomly alternated between the conditions, with stipulation that no condition and type repeated successively. After every 2-3 blocks, participants were able to take a break (to eat some snacks, stretch while seated, etc.) as a physical and mental reprieve to protect against fatigue (10 breaks overall, average duration 1-2 minutes). The EEG net was not removed during breaks although no EEG data were recorded. The total length of experiment, including breaks, was approximately 40-45 minutes.

Electrophysiological Recording and Analysis

Participants sat while the experimenter applied and adjusted the EEG sensor net. Participants were given task instructions (described above) and EEG recording began. Parents were allowed to remain in the room with their child during recording, and both parent and child were asked to be as quiet as possible throughout the procedure. Net placement, adjustment, and instructions took approximately 15 minutes.

EEG was recorded continuously from scalp electrodes using the Geodesic Sensor Net (Tucker, 1993), a network of 128 Ag/AgCl electrodes embedded in an elastic geodesic tension structure. Impedance for all electrodes was kept below 50 KΩ (this EEG system used high-impedance amplifiers, thus the relatively high electrode impedances), and all recordings were referenced to the vertex (Cz). Signals were amplified with a 0.1 Hz to 100 Hz elliptical bandpass filter and digitized at 250 Hz sampling rate. Continuous EEG data were segmented to epochs of 1500 ms after stimulus onset with a 100 ms pre-stimulus baseline.

Artifacts were identified in the EEG via the following steps. For each trial, channels were marked for artifact if a running average of activity exceeded 40 μV (detecting sharp transitions in the signal). Subsequent to this automated process, each trial was manually inspected. Trials with more than 15 channels with artifact were excluded. For trials with less than 15 channels with artifact, an algorithm that derives values from neighboring channels via spherical spline interpolation replaced bad channels. EEG data were corrected for eye-blink and eye-movement artifacts using the Gratton, Coles, and Donchin (1983) algorithm. EEG data were re-referenced off-line against the average reference. Epochs of EEG data in the same condition were averaged to derive the ERP data. An average of 79 usable trials total (no less than 15 usable trials per condition) per participant were used. This average total is lower compared to the average total useable trials in Bowman et al. (2012). The difference is likely because the ERP task was easier for the older children in the present study and therefore much less compelling; thus children in the present study became restless more quickly and had a more difficult time sitting still through the large number of ERP trials compared to the younger children who were more captivated by the task. Importantly though, for the present study, the average total usable trials was equivalent across all three conditions (27, 25, and 25 trials for beliefs, desires, and physical conditions respectively) ensuring that all conditions were represented equally in the ERP data, and our cutoff of no less than 15 useable segments per condition is identical to the cutoff used in Bowman et al. Prior to analysis, the ERP data were corrected to the 100 ms pre-stimulus baseline and digitally filtered with a 30 Hz low-pass filter.

Results

The present study examined belief- and desire-reasoning in 10- and 11-year-old children, and employed the same stimuli and methodology used in a parallel ERP investigation of younger children ages 7- and 8-years-old (Bowman et al., 2012) to provide a developmental comparison. Thus, results from the present study are presented separately but also in conjunction with the 7- and 8-year-old data and analyses from Bowman et al. where appropriate.

Preliminary Analyses

Given the exclusions in our sample, we confirmed that included and excluded children did not differ on demographic or performance variables – pairwise *t*-tests indicated these groups were equivalent (all *p*s > .12) on age, gender, and task performance variables (mean accuracy and mean reaction time for both memory control and target questions). Thus, our target group is representative of the entire sample.

Behavioral Performance Accuracy

Children at this age were significantly and substantially correct at solving all three types of problems: physical control tasks (91.5%), diverse-desires tasks (91.9% correct) as well as diverse-beliefs tasks (80.7%). As expected based on results from Bowman et al. (2012) and previous behavioral research, however, children were better at solving diverse-desires and physical tasks compared to diverse-beliefs: *t*(19) = -3.28, *p* = .004; *t*(19) = -2.88, *p* = .010 for Beliefs condition versus Desires and Physical conditions, respectively. Desires and Physical conditions did not differ from each other: *t*(19) = .27, *p* = .789. As shown in Table 5.1, this overall pattern is consistent for both the 10- and 11-year-old data and the 7- and 8-year-old data, and is in line with behavioral research demonstrating that, compared to desire-understanding, belief-understanding emerges later in early childhood (see meta-analysis in Wellman & Liu, 2004), and can be less accurate/fluent in older children and adults (e.g., Malle, 2004).

At the same time, when comparing performance accuracy between the present study's 10- and 11-year-old sample and the 7- and 8-year-old sample from Bowman et al. (2012), there is a clear developmental effect. Accuracy for the focal Beliefs and Desires conditions was greater in the older sample compared to the younger sample (see Table 5.1 for means, standard deviations and t-test results for condition accuracy in the two age groups). Moreover, across the two age groups, age positively correlated with accuracy for both Beliefs and Desires (see Table 5.2), demonstrating a clear developmental effect of improvement for belief- and desire-reasoning as children age over middle to late childhood.

As also depicted in Table 5.2, accuracy across the three conditions was inter-related; in particular, Desires accuracy positively correlated with both Beliefs and Physical accuracy. Given these inter-relationships, it is possible that the improvements in belief- and desire-reasoning described above could be reflecting domain-general improvements in cognitive functioning (e.g., working memory, speed of processing).

The present study contains no direct measures of working memory or speed of processing, but it seems clear that such factors should equally affect improvement in all three conditions, including performance in the Physical condition (which involved the same memory and perceptual task demands but did not require mental-state reasoning), rather than affecting improvements in mental-state understanding specifically. Thus, to address this issue indirectly and to further examine relations between performance accuracy and age specifically, I conducted three stepwise regressions each with accuracy for a given condition (e.g. Beliefs accuracy or Desire accuracy) as the dependent variable, and age and accuracy for the remaining two conditions as predictors. I reasoned, in particular, that if the correlation between age and performance in the Beliefs condition is a reflection of development in belief-reasoning specifically, then we would expect performance accuracy in the Beliefs condition to be best predicted by age, beyond any predictive effects of the Physical control condition or the Desires condition.

Results suggest advancement in belief-reasoning specifically, as children age. As clear in Table 5.3, age was the sole best predictor of Beliefs accuracy ($F(1,34) = 9.45$, $p = .004$); whereas neither Physical accuracy nor Desires accuracy significantly added to the model. These results confirm that, independent of domain general advancements that are also likely occurring over this time (indexed in this case by performance in the Physical condition), there are advancements in belief-reasoning as children develop from middle to late childhood. In contrast, age did not significantly predict either Desires accuracy or Physical accuracy (see Table 5.3). Rather, Desires accuracy was best predicted by Physical accuracy ($F(1,34) = 20.45$, $p < .001$), and Physical accuracy was best predicted by Desires accuracy ($F(1,34) = 20.45$, $p < .001$). Thus, from middle to late childhood, results suggest that developments in theory-of-mind reasoning concern developments in belief-reasoning specifically, beyond developments in domain general cognitive skills, and beyond developments in desire-reasoning.

ERP Data Analyses

The analytic approach for ERP data analysis follows that of Bowman et al. (2012). In addition, findings from that younger 7- and 8-year-old sample guide analyses in the present study in order to directly explore any developmental changes between the younger and older child samples that might be present in the ERP data.

The lesser accuracy for Beliefs conditions compared to Desires and Physical conditions in the present study's 10- and 11-year-old sample described in the section above presents an issue for ERP data analysis. Bowman et al. (2012) dealt with this same issue by focusing especially on ERP patterns for correct trials (correct-trial analysis). This approach is standard in developmental neuroscientific investigations and is an attempt to better equate neural activations for the same tasks across age (e.g., Casey, Giedd & Thomas, 2000). For the present older sample in particular, incorrect performance trials likely represent a high proportion of inaccuracies due to lack of attention and focus (rather than focused yet incorrect reasoning). Thus, for the present study, I report only results of the correct-trial analysis. (And, indeed, full-trial analyses of the present 10- and 11-year-old data yielded ERP waveforms that were noisier compared to the correct-trial data, and difficult to interpret.)

Due to the reduced number of correct trials overall, five participants did not have at least 15 artifact-free correct trials per condition, and thus were excluded from correct-trial analyses. The final sample for correct-trial analyses was 15 children (8 males; age range = 120-142 months; *M* = 130 months, *SD* = 8.02). Importantly, this smaller sample did not differ from the original sample of 20 on age, gender, accuracy, or reaction time (all *p*s > .49).

The Physical condition was designed as a control for the two mental state conditions. Thus, differences in waveforms between Beliefs and Physical or between Desires and Physical conditions reveal components associated with reasoning about *mental states* over and above common components for processing these parallel task formats and making these comparative judgments; whereas differences between the Beliefs and Desires conditions reveal differences in processes for reasoning about beliefs versus desires.

As a first step, difference waves were calculated by subtracting the mean amplitude in one condition from the mean amplitude in another condition (i.e., Beliefs minus Physical, Desires minus Physical, and Beliefs minus Desires), at each of the 128 channels. Topographical maps of the difference waves show a clear difference between both Belief and Physical conditions and Desire and Physical conditions concentrated in mid-frontal (and right-frontal) scalp regions (see Figure 5.3). These differences appear early around 200 ms post-stimulus onset, but diminish after 800 ms post-stimulus. In contrast, there is no apparent difference between Beliefs and Desires conditions across the time course. To confirm the condition differences revealed in the topographic maps, we conducted two analyses: analysis of a subset of channels systematically encompassing locations from left to right and top to bottom over the scalp (subset analysis), and a more focused region of interest (ROI) analysis.

Subset analysis. Following a common analytic approach to avoid 128 multiple comparisons (e.g., Bowman et al., 2012; Liu et al., 2009a), we selected a grid of channels from the larger group of 128 channels to systematically sample locations from left to right (laterality) and from anterior to posterior (caudality) over the scalp. Figure 5.4 displays the grand average waveforms for all three conditions from all channels in this 5x5 grid. Visual inspection of the waveforms in Figure 5.4 matches the pattern displayed in the topographic maps: there is a clear difference between the Physical condition and both mental-state conditions (more negative), concentrated in the mid-frontal and right-frontal channels (e.g., channels Fz and F2). Again, this difference appears early around 200 ms, and remains through 500 ms, but diminishes in later epochs beyond 800 ms post-stimulus. I examined these differences through analyses of the mean amplitude of the ERP data for all 25 channels in the grid within five different epochs: 200-250 ms, 350-500 ms, 600-800 ms, 800-850 ms, and 1100-1400 ms post-stimulus onset. Each epoch represents a distinct portion of the waveform (i.e., peak, gradual increase in slope, attenuation of slope, horizontal slope, and tail-end of waveform, respectively). Importantly, these epochs closely match the epochs used in Bowman et al. (2012), and the 25 channels selected for analyses are identical to those selected for the subset analysis in Bowman et al., to allow comparison of ERP waveforms across the two age-groups. When necessary, *p*-values were adjusted using the Greenhouse-Geisser correction – a common adjustment in most ERP research (e.g., Bowman et al., 20120; Liu et al., 2009a; Van der Cruyssen, Van Duynslaeger, Cortoos, & Van Overwalle, 2009). Alpha was set at $p < .05$.

Five separate omnibus 3 (condition) x 5 (laterality) x 5 (caudality) repeated-measures analyses of variance (ANOVAs) (one for each of the five epochs) were conducted to compare the mean amplitude across all three conditions. We focus on main effects of condition as well as three-way condition x laterality x caudality interactions because our focal questions concern differences in ERPs across *conditions* and where these condition differences concentrate on the scalp. Main effects of caudality or laterality and 2-way interactions (subsumed by the focal three-way patterns) are not considered further.

Results of the omnibus ANOVAs yielded significant condition effects in the 200-250 and 350-500 ms epochs: $F(2,28) = 25.60, p < .001$; $F(2,28) = 7.10, p = .005$, respectively. The 200-250 epoch also demonstrated a significant condition x laterality x caudality effect: $F(32,448) = 2.63, p = .036$. No condition effects were found in the later epochs of 600-800, 800-850, and 1100-1400 ms (*p*s > .280) for these analyses. To further examine the condition differences evidenced in the 200-250 and 350-500 ms epochs, three 2 (condition) x 5 (laterality) x 5 (caudality) repeated measures ANOVAs on the mean amplitude in each of the two epochs directly compared 1) Desires versus Physical, 2) Beliefs versus Physical, and 3) Beliefs versus Desires conditions. As before, we focus on main effects of condition, and condition x laterality x caudality effects.

Both Beliefs-Physical and Desires-Physical comparisons yielded main effects of condition in the 200-250 ms epoch: $F(1,14) = 32.03, p < .001$; $F(1,14) = 55.89, p < .001$, respectively. This same pattern of results was true for the 350-500 ms epoch as well: $F(1,14) = 8.47, p = .011$; $F(1,14) = 9.45, p = .008$, for Beliefs-Physical and Desires-Physical comparisons, respectively. The 200-250 ms epoch also yielded a significant condition x laterality x caudality interaction for the Desires versus Physical comparison: *F*(16,224) = 4.98, *p* = .003. In contrast, comparisons of Beliefs and Desires conditions showed no significant condition effects and no significant condition x laterality x caudality interactions in either epoch (all *p*s > .22). These results indicate that the neural activations for beliefs and desires differentiate equally from the physical control condition, but do not differentiate from each other, demonstrating a 'mental-versus-physical' distinction in these early epochs in frontal channels. The significant interaction effect in the 200-250 ms epoch suggests that for the desires-versus physical effect, this distinction is more diffuse, covering a wider range of scalp locations.

This pattern of results parallels what is depicted in Figure 5.3 and Figure 5.4. That is, these figures show a clear difference between Beliefs versus Physical activations and Desires versus Physical activations but no difference between Belief versus Desire activation patterns, concentrated in the mid-frontal (and initially also in right mid-frontal) scalp regions, and evident early in the ERP time course but diminishing later. Moreover, in the 200-250 ms epoch, the strongest Desires-Physical distinctions are more diffuse compared to the Beliefs-Physical distinction (see Figure 5.3), in line with the condition x caudality x laterality interaction effect demonstrated for Desires versus Physical (but not for Beliefs versus Physical) in this earliest epoch. Thus, results from the ANOVAs, in conjunction with results from the topographical maps and ERP waveforms, provide clear evidence for a mental-versus-physical distinction in mid-frontal and right-frontal scalp regions.

Frontal ROI analysis. Following from Bowman et al. (2012), I conducted an ROI analysis to further examine these frontal effects, and confirm that they encompassed a cluster of channels. In line with Bowman et al. and other ERP investigations (e.g., Sabbagh, Moulson, & Harkness, 2004), cluster effects were considered significant only when condition differences occurred in at least four adjacent channels, and each channel showed a similar pattern of activity that achieved significance or near significance. We selected five channels in the mid-frontal scalp to represent our mid-frontal cluster. Importantly, this 'mid-frontal' region of interest was selected based on two sets of prior information: 1) existing neuroscientific research with 10- and 11-year-old children consistently implicates the mid-frontal region of the MPFC as involved in theory-of-mind reasoning (e.g., Sabbagh et al., 2009; Sommer et al., 2010; Saxe et al., 2009), and 2) most specifically, when the present study's ERP task was administered to adults (Liu et al., 2009a) and children 7- and 8-years old (Bowman et al.), there was a clear 'mental-versus-physical' effect in mid-frontal and right mid-frontal scalp regions. Moreover, mid-frontal channel Fz from the 5x5 grid (Figure 5.4) most strongly displays the mental-versus-physical effect, and the topographical maps (Figure 5.3) show that, beyond 250 ms, this effect concentrates in further mid-anterior regions. Thus, channel Fz along with four other neighboring anterior channels (EGI channels 14, 15, 16, and 10) constituted our 5-channel *mid-frontal* cluster for this ROI analysis (see Figure 5.5a).

As described earlier, the 7- and 8-year-old sample in Bowman et al. also yielded a clear mental-versus-physical distinction in mid-frontal scalp regions. A similar 5-channel cluster ROI analysis confirmed the effect in that sample. However the cluster selected in Bowman et al. had a scalp distribution that spread more to the right side of the head, encompassing the *right* mid-frontal scalp (see Figure 5.5b). Thus, for the present study's 10 and 11-year-old sample, we selected an additional *right mid-frontal* cluster (see Figure

5.5c) for analysis, using the same 5 channels used in Bowman et al. (channel FC2 and EGI channels 112, 113, 118, and 124), to compare any differences in the right mid-frontal mental-versus-physical effect that may have occurred between the two age groups. Moreover, the mental-versus-physical distinction in Bowman et al. began early in the 200-250 ms epoch, but was sustained through 850 ms. Indeed, visual inspection of the topographical maps and ERP waveforms in Figures 5.3 and 5.5a for the present study data suggest that the mental-versus-physical effect extends beyond 500 ms. Thus we examined condition differences in the 200-250 and 350-500 ms epochs, as well as in the 600-800 and 800-850 ms epochs (the same epochs used in Bowman et al.) to determine the extent of the mental-versus-physical condition effect over the ERP time course.

Thus, for both mid-frontal and right mid-frontal clusters, three 2 (condition) x 5 (channel) repeated measures ANOVAs on the mean amplitude in the 200-250, 350-500, 600-800, and 800-850 ms epochs directly compared Desires-Physical, Beliefs-Physical, and Beliefs-Desires conditions. Main effects of condition (and not condition x channel interactions) are focal because our question concerns the extent to which condition effects are observed within our channel *cluster*.

Mid-frontal cluster. For the mid-frontal cluster, Beliefs-Physical and Desires-Physical comparisons yielded main effects of condition in the 200-250 and 350-500 ms epochs (see top panel of Table 5.4). The Desires-Physical comparison also yielded a significant condition effect in the 600-800 ms epoch, but diminished in the 800-850 ms epoch; no significant condition effects were found for the Beliefs-Physical comparison beyond 500 ms (top panel Table 5.4). Beliefs-Desires comparisons showed no significant condition effects in any of the epochs.

Paired-samples *t*-tests confirmed that these patterns of results occurred in each of the five channels within the cluster. Beliefs and Desires conditions did not differ from each other in any of the five channels within the cluster, in any of the four epochs (all *p*s > .12). In contrast, for all five channels, both Beliefs and Desires differed significantly from the Physical condition in the 200-250 and 350-500 ms epoch (all *p*s < .04). The Desires-Physical distinction extended into the 600-800 ms epoch in all five channels as well (all *p*s < .03), though it diminished in all five channels in the 800-850 ms epoch. In this latest epoch, the Desire-Physical difference reached neither significance nor marginal significance in EGI channels 10 and 16, (*p*s > .13), and reached only marginal significance in channel Fz and EGI channels 14 and 15 (*ps* < .09).

Right mid-frontal cluster. Results for the right mid-frontal cluster (identical to the cluster selected in Bowman et al., 2012) were similar to those of the mid-frontal cluster for the earliest epoch of 200-250 ms: Beliefs and Desires activations did not differ from each other in either the overall cluster ANOVA (see bottom panel of Table 5.4) or in any of the individual channel *t*-tests (all *p*s > .16); whereas both Beliefs-Physical and Desires-Physical comparisons yielded significant condition effects in the overall cluster (bottom panel Table 5.4) and in all five individual channels (all *ps* < .009).

But in contrast to the mid-frontal cluster results, the pattern of effects for the *right* mid-frontal cluster diminished more quickly in the ERP time course, especially for the Beliefs-Physical comparison. Specifically, in the 350-500 ms epoch, ANOVAs revealed a significant condition effect for the Desires-Physical comparison, but only a marginally significant effect for the Beliefs-Physical comparison. And by the 600-800 ms epoch neither Beliefs-Physical nor Desires-Physical comparisons yielded significant condition effects (bottom panel Table 5.4).

Individual channel *t*-tests further confirmed this pattern of diminished distinctions in later epochs. For the Beliefs-Physical distinctions, in the 350-500 ms epoch, *t*-tests yielded significant differences in amplitude in only 2/5 channels (channel FC2 and EGI channel 124) (*p*s < .05), with EGI channels 113 and 112 showing a marginally significant difference (*p*s < .080) and EGI channel 118 showing no significant difference (*p* = .11). As seen in Figure 5.5c, the channels that show the most diminished effect are located in the most right-of-center, and most posterior positions of the right mid-frontal cluster.

For the Desires-Physical comparison, *t*-tests revealed significant condition differences in 4/5 channels in the 350-500 ms epoch (channel FC2 and EGI channels 113, 118, and 112) (*p*s < .05), with EGI channel 124 reaching marginal significance (*p* = .080). And for the 600-800 epoch, the Desires-Physical distinction reduced to marginal significance in 3/5 channels (channel FC2 and EGI channels 124 and 112) (*ps* < .086), with EGI channels 113 and 118 channels showing no significant difference at all (*ps* > .12). Here too, the channels that show the most diminished effect for the Desires-Physical comparison are located in the most right-of-center, and most posterior positions of the cluster (see Figure 5.5c).

Results of these frontal ROI cluster analyses converge directly with the earlier analyses, and further confirm what is shown clearly in Figures 5.3, 5.4, and 5.5a and c: in mid-frontal scalp regions, Beliefs and Desires conditions differ equally from the Physical control condition but do not differ from each other (with mental conditions both more negative compared to the physical condition). These effects begin early in the ERP time course and diminish by later epochs.

The mid-frontal pattern of effects is comparable to the ERP findings in the parallel 7- and 8-year-old sample in Bowman et al. (2012): those younger children also demonstrated a mental-versus-physical distinction, concentrated in right mid-frontal scalp regions, with belief and desire activations equally more negative compared to activations for the physical condition but not different from each other (see Figure 5.5b). However, results from the present study also reveal two intriguing differences from the younger sample data. First, though still present in right mid-frontal scalp regions (as in the younger age group), the most robust mental-versus-physical effect for our older age group was shifted away from the right side of the scalp towards the anterior mid-frontal regions. Second, the mental-versus physical effect in the younger sample extended late into the ERP time course, showing mental activation significantly more negative than physical activation up through 850 ms. In contrast, in our older sample of children, the mental activation was significantly more negative compared to physical activation *only* in the earlier epochs; later in the time course this mental-physical distinction diminished. See Figure 5.5 for global frontal comparisons between the 10- and 11-year-old and 7- and 8-year-old samples.

Right-posterior ROI analysis. As outlined earlier, there are *a priori* reasons for closely examining Beliefs versus Desires differences in right posterior regions. Similar to the frontal ROI analysis, a close examination of belief versus desire activations in a right posterior ROI is warranted from two sets of prior information: 1) existing neuroscientific research with 10- and 11-year-old children consistently implicates the right posterior region of the right TPJ as involved in theory-of-mind reasoning (e.g., Saxe et al., 2009; Gweon et al., in press), and 2) most specifically, when the present study's ERP task was administered to adults (Liu et al., 2009a) and children 7- and 8-years-old (Bowman et al.), there was a clear 'beliefs-versus-desires' distinction in right posterior scalp regions. Moreover, as can be seen in Figure 5.4 and still more clearly in Figure 5.6a, visual inspection of the grand average waveforms for the right posterior channel CP6 in our data suggests greater mean amplitude for Beliefs compared to Desires. This channel was the same focal channel that showed the beliefs-greater-than-desires effect in the younger sample in Bowman et al (see Figure 5.6b). Thus, we selected channel CP6 along with 3 additional neighboring channels showing a similar effect (channel P6 and EGI channels 98, and 94) to serve as a right posterior 4-channel cluster (Figure 5.6a). As evident in Figure 5.4 and still more clearly in Figure 5.6a, the belief-desires distinction begins in the later epochs beyond 1100 ms post stimulus, and indeed in Bowman et al. a similar effect was demonstrated in the tail end of the waveform as well (see Figure 5.6b). Thus, I examined belief and desire activations in both a 1100-1200 ms epoch and a 1300-1400 ms epoch. In Bowman et al., the beliefs-versus-desires effect was also demonstrated earlier in the 600-800 ms epoch (Figure 5.6b). Though visual inspection of our data suggests a beliefs-desires distinction is likely not evident earlier in the time course, I examined belief and desire activations in the 600-800 ms epoch as well to most closely parallel analyses in Bowman et al.

Three separate 2 (condition) x 4 (channel) repeated measures ANOVAs compared mean amplitude in Beliefs versus Desires conditions in the 600-800, 1100-1200, and 1300-1400 ms epochs. Neither the 600-800 nor the 1100-1200 ms epochs yielded significant condition effects (see Table 5.5). There was, however, a significant beliefs-versus-desires condition effect in the 1300-1400 ms epoch (Table 5.5). Paired-samples *t*-tests comparing mean Beliefs and Desires amplitude in each of the 4 channels yielded significant results in 2/4 channels (channel CP6 and EGI channel 98 (*p*s < .05) and marginally significant results in the remaining two channels (channel P6 and EGI channel 94) (*p*s < .084), meeting criteria for a significant cluster effect overall in this late epoch. In contrast, repeated measures ANOVAs for an equivalent cluster of four *left* posterior channels showed no significant difference between Beliefs and Desires conditions in any of the four epochs examined (all *p*s > .22). These results thus confirm the presence of a beliefs-versus-desires distinction (with belief activation more positive compared to desire activation) specific to right posterior scalp regions, occurring late in the time course. This distinction parallels the beliefs-desires distinction found in the 7- and 8-year-old sample in Bowman et al. (2012), which also occurred in right posterior scalp regions (including channel CP6), and which was evident late in the ERP time course (see Figure 5.6b). In this younger age group however, the beliefs-desires distinction was also evident earlier in the time course in the 600-800 ms epoch. Thus, in contrast to findings from Bowman et al., the difference in neural activation for beliefs- versus desire-reasoning in the present study's older sample was evident only in the latest portion of the ERP time course, demonstrating a reduction of a beliefs-desires distinction in earlier epochs as children age.

Discussion

The present study investigated the neural correlates of belief- and desire-reasoning in 10- and 11-year-old children. We found evidence for two distinct neural patterns associated with reasoning about mental states. First, frontal neural activations for both mental-state conditions (Beliefs and Desires) differentiated from the Physical control, but did not differentiate from each other. These effects were demonstrated by significantly more negative activations for the mental conditions compared to physical activations, they occurred most robustly in the earlier epochs of the ERP time course from 200-500 ms post stimulus, and they were located most robustly in the mid-frontal scalp. Second, the neural activation associated with belief-reasoning differentiated from activation for desire-reasoning. This effect was located in the right-posterior scalp and was demonstrated by significantly more positive activation for the beliefs condition compared to desire activation. It occurred late in the ERP time course from 1300-1400 ms post stimulus.

Broadly, findings from the present study parallel those from Bowman et al. (2012) investigating 7- and 8-year-old children, as well as those from Liu et al. (2009a) investigating adults. Specifically, results from all these studies show that, for both children and adults, two distinct patterns of neural activation exist: one pattern in mid-frontal regions in which both beliefs and desires differentiated from the physical control but not from each other, and a second pattern in right posterior regions in which activation for belief-reasoning separated from desire activations. In advance of the present study, only Bowman et al. (2012) had shown that two such neural systems for mental-state reasoning were present even in childhood. But in total, these data now present strong evidence for two neural systems for mental-state reasoning: a common system for reasoning about mental-states more generally, including beliefs *and* desires (in neural regions associated with the mid-frontal scalp), and a separate system (located in right posterior scalp regions) that potentially supports belief-reasoning specifically—distinct from desire-reasoning.

Taken together, these data suggest an important explanation for what may, at least in part, underlie the progression from an early understanding of desires to a later understanding of beliefs—a development first and strongly demonstrated in behavioral data from preschoolers (e.g., Wellman & Liu, 2004). Specifically, an understanding of beliefs may build on prior understanding of desires as belief-reasoning first appears and then becomes more distinct and accurate—evidenced by additional neural substrates (in right posterior regions) recruited for specifically belief-reasoning, beyond a common neural system (in mid frontal regions) for reasoning about mental-states more generally. Behavioral performance accuracy from Bowman et al. (2012), Liu et al. (2009a), and the present study are in line with this possibility. Adults (who achieved near-perfect accuracy) and our older 10- and 11-year-old sample (who exhibited much improved accuracy compared to the younger 7- and 8-year-olds, particularly for belief-reasoning) demonstrated clearly distinct activation for belief-reasoning compared to desire-reasoning in right posterior scalp regions. However, the younger 7- and 8-year-old children (who struggled considerably more with belief-reasoning tasks compared to both adults and our older child sample) also demonstrated a right posterior belief-desires distinction, but *only* when neural activations reflected solely accurate reasoning. These results thus lay foundation for future research to examine children even younger than 7 years, in order to investigate whether a second system for belief-reasoning (as distinct from desire-reasoning) exists even in the preschool years, when belief-reasoning, although beginning to emerge, is even less accurate, and is undergoing much stark behavioral advancement.

Results from the present study provide important developmental comparisons to both the younger 7- and 8-year-old children in Bowman et al. (2012) and to the adults in Liu et al. (2009a), providing more fully developmental data. The data and resulting comparisons are especially informative because all three of these studies used the same ERP stimuli and procedures. Thus, the key results from the present study are discussed separately but also in conjunction with those adult and younger child findings. To begin, I discuss further the nature and implications of the developments in behavioral accuracy apparent across these ages. Then, I turn more focally to the ERP data. This ERP discussion returns to several of the issues outlined in the introduction, including age-related differences in timing, polarity, and location of ERP effects.

Behavioral Accuracy Advancements

Behavioral accuracy for the ERP task (which required multiple target judgments across each of the three conditions) increased across age from 7 years to adulthood. More specifically, when behavioral performance accuracy for belief-, desire-, and physical-reasoning was compared across middle and later childhood from 7- to 11-year-olds, there was a clear developmental advancement in specifically belief-reasoning. Regression analyses showed that age was the sole best predictor of belief-reasoning; neither desire-reasoning nor physical reasoning significantly contributed to the model. Moreover age did not predict either desire-reasoning or physical-reasoning. Given the similarities in domain-general task demands (e.g., working memory, attention, perceptual processing) across all conditions, these results also provide evidence that belief-reasoning advancements occur beyond advancements in domain general skills (at least as indexed by performance on physical- and desire-reasoning tasks) that are likely also occurring over this time period.

Of course future research should include direct measures of executive functioning, and IQ in order to directly confirm this conclusion. Yet, these behavioral results are in line with numerous other behavioral findings demonstrating that, in early childhood, belief-reasoning is more difficult compared to desire-reasoning, and develops later (e.g., Wellman & Liu, 2004). Critically, our results demonstrate that a similar pattern continues to occur even later in childhood from 7- to 11-years-old, two to six years beyond the age at which children reach ceiling on theory-of-mind judgments when given single, clear standard tasks. Results are also in line with the few existing behavioral studies that examine belief and general mental-state reasoning in older children and adults (e.g., Keysar et al., 2003, Apperly et al., 2011). However the crucial addition of the desire-reasoning contrast in the present study further suggests that these later theory-of-mind developments are likely driven by advancements in belief-reasoning specifically, rather than advancements in general mental-state reasoning or advancements of other types of mental-state reasoning such as desires.

ERP Components

As background for these comparisons, recall our ERP task was designed such that all three conditions (Beliefs, Desires, and Physical) had the same perceptual and linguistic structure including similar two-part comparisons. Following standard neuroimaging subtraction methodology then, any differences in neural activation across conditions can be attributed to differences in reasoning about the *content* of each condition (i.e., belief-reasoning, desire-reasoning, or reasoning about physical locations) beyond the memory and processing demands common in all conditions. Similarly, comparison of these condition differences across age are likely to reflect development in reasoning specific to mental states rather than information processing capacities alone.

Frontal effect: Distinct activations for mental- versus physical-reasoning. As expected, in mid-frontal scalp regions, children exhibited a pattern of neural processing for reasoning about diverse-beliefs and diverse-desires that was separate from the neural pattern for the physical control condition. As noted, this result parallels both the adult ERP findings in Liu et al. (2009a) and the 7- and 8-year-old findings in Bowman et al. (2012) in which both adults and younger children showed overlapping activation for belief- and desire-reasoning, with both mental-state conditions differing equally from activation for the physical control in the mid-frontal and right mid-frontal scalp. More generally, these results are in line with existing research demonstrating that mental-state reasoning recruits specific neural substrates in medial frontal regions of the brain in both adults and children (e.g., Carrington & Bailey, 2009; Sommer et al., 2010). These substrates have been shown to be associated with mental-state reasoning even after controlling for domain general skills (i.e., executive functioning; Sabbagh et al., 2009, for 4-year-olds), and they have been shown to be distinct from the substrates supporting non-mental processing (i.e., physical reasoning; Saxe et al., 2009; Gweon et al., in press, for 6- through 11-year-olds). Thus, our results follow the common pattern of specific neural substrates supporting mental-state reasoning, and add increasing evidence that such specificities exist even at young ages. Moreover, they suggest a common neural system for reasoning about both beliefs *and* desires, in mid-frontal scalp regions.

Age-related differences in location, timing, and polarity of the mental-versus-physical effect. Beyond the overall similarities with both the adult (Liu et al., 2009a) and younger child (Bowman et al., 2012) samples, results of the present study also demonstrate additional similarities to and differences from the ERP activation patterns in these other two age groups in terms of location of the effect, and the timing and polarity of the effect. Indeed, considering all three age groups together, ERP data from the present study suggest that neural correlates for belief- and desire-reasoning in 10- and 11-year-olds may represent a transitional state in the organization and properties of the neural substrates supporting these two types of mental-state reasoning.

To illustrate, recall the comparisons between the adult ERP data from Liu et al. (2009a) and the 7- and 8-year-old child data from Bowman et al. (2012), summarized in Figure 5.1. For adults, the mental-versus-physical distinction was demonstrated by more positive belief and desire activations compared to physical activations. This distinction was located in the mid-frontal scalp, captured most focally by mid-frontal channel Fz, and it occurred later in the ERP time course beyond 800 ms. For 7- and 8-year-old children (Bowman et al.), a similar mental-versus physical distinction was demonstrated in mid-frontal regions, though the effect was more diffuse, spreading to *right* mid-frontal scalp regions. For these younger children, the effect was also demonstrated beyond 800 ms post stimulus, but the distinction began early in the time course at 200 ms post stimulus and moreover was demonstrated by more *negative* belief and desire activations compared to physical activations.

The present study's frontal effect shows both remnants of the effect as demonstrated in 7- and 8-year-olds as well as a shift toward the particular components of the effect as demonstrated in adults. Specifically, similar to the younger age group, children in the present study also exhibited a mental-versus-physical distinction early in the ERP time course beginning at 200 ms post stimulus and sustained through 500 ms[6]. Moreover, the effect occurred in right-mid frontal scalp regions, including in the same focal channel FC2. However, two intriguing differences emerged. First, though still present in right mid-frontal scalp regions (as in the younger age group), the most robust mental-versus-physical effect for the older age group was shifted away from the right side of the scalp towards the anterior mid-frontal regions. Second, for the younger sample, the mental-versus physical effect extended late into the ERP time course, showing mental activation significantly more negative than physical activation up through 850 ms. In contrast, in our older sample of children, the mental activation was significantly more negative compared to physical activation *only* in the earlier epochs; later in the time course this mental-physical distinction diminished.

Crucially, where the present study and the younger sample findings diverge, the present study and the *adult* sample begin to *converge*. Indeed, the most robust mental-versus-physical effect in our 10- to 11-year-old sample was concentrated in the mid-frontal scalp, including in focal channel Fz—the same focal channel that demonstrated the mid-frontal effect in the adult sample (from Liu et al. 2009a). Moreover, the effect

---

[6] As reasoned in Bowman et al. (2012), this early distinction possibly represents mental-state decoding (on-line attribution of mental-states to individuals; e.g., Tager-Flusberg & Sullivan, 2000; Pineda & Hecht, 2009; Sabbagh, 2004) that was elicited during the information phase of the trial but that was still being processed in children by the time the ERP mental-state reasoning event (in which representations of individuals' mental-states are used in order to predict and make judgments about actions) was recorded. In contrast, adults likely completed the decoding process earlier due to more efficient processing in general compared to children (e.g., Kail, 1991), and thus did not show an early mental-versus-physical distinction in the ERP waveform.

was also less diffuse, occurring most robustly in fewer channels compared to the younger 7- to 8-year-olds, showing another similarity to the adult data as well as a pattern in line with neuroscientific research that demonstrates a common shift from more diffuse to more localized effects with age (e.g., Casey et al., 2000).

As captured in Figure 5.1, in the adult sample, the mental-versus-physical frontal effect occurred later in the ERP time course and was demonstrated by more positive belief and desire activations compared to physical activations. In the present 10- and 11-year-old sample, the mental-versus-physical effect was demonstrated in earlier epochs by significantly more negative beliefs and desires activations compared to the physical condition, but this distinction had diminished by later epochs. Speculatively, this shift from more negative mental activations compared to physical activations to more equal activations with the physical condition in these later epochs could be an indication of a shift in polarity as children age. That is, it is possible that our 10- and 11-year-old data capture the beginning of a transition from a mental-*less-than*-physical mid-frontal effect (demonstrated in the 7- and 8-year-old sample), to a mental-*greater-than*-physical mid-frontal effect (demonstrated in the adult sample), and thus characterizes an age-related change in polarity of the underlying neuronal dipoles that support belief- and desire-reasoning.

This proposal is speculative in two fashions. First, descriptively, investigations of belief and desire neural correlates in children older than age 11 years would be required to show increasingly more positive mid-frontal mental (versus physical) activations later in the ERP time course in order to shed more light on a possible shift in polarity. Second, the meaning of polarity in ERP findings, and in particular *developmental* shifts in polarity, are little discussed or understood. Indeed, there is no research that we know of that directly examines links between changing polarity and reorganization of underlying neural substrates. Nonetheless, it is possible that for the mental-versus-physical effect, both the shift from right mid-frontal to mid-frontal locations and the potential later shift in polarity indicate underlying reorganization/development of neural substrates supporting belief- and desire-reasoning. Indeed, several researchers suggest that age-related changes in ERP components are linked to underlying neural reorganization and developments in cognitive reasoning (e.g., Marshall, Drummey, Fox & Newcombe, 2002; Johnson & de Haan, 2001).

Existing fMRI investigations of theory-of-mind development provide some indirect support for the general hypothesis that shifts in ERP components reflect development and reorganization of neural systems. Specifically, several fMRI studies of children 6- through 11-years-old implicate the medial prefrontal cortex as recruited for theory-of-mind reasoning at a greater magnitude compared to adult samples (e.g., Pfeifer et al., 2009, Sommer et al., 2010). However, beyond age 11 years, this increased prominence in MPFC recruitment is reduced (e.g., Moor et al., 2012). Conceivably, our 10- and 11-year-old data are capturing the beginning of such a reduction in mid-frontal regions as children age beyond 11 years, demonstrated by a reduction of a mental-versus-physical distinction later in the ERP time course. Future research that tracks the development of ERP activation patterns across age will be critical in shedding light on how the timing, polarity, and location differences evidenced in the present study relate to both behavioral and neural developments in theory-of-mind reasoning.

Right posterior effect: Distinct activations for belief- versus desire-reasoning. A key finding in the present study was the pattern of neural processing specific to belief-reasoning—distinct from desire-reasoning—concentrated in right posterior scalp regions. Specifically, activation for the beliefs condition was significantly more positive compared to desires activations—a distinction that occurred late in the ERP time course beyond 1300 ms post stimulus. These results are also in line with results from the parallel adult and younger child ERP samples from Liu et al. (2009a) and Bowman et al. (2012), in which both adults and 7- and 8-year-old children showed more positive belief activation compared to desire activation in right posterior scalp. The location of the right posterior effect in the present study overlapped with the location of the effect in both younger child and adult samples: In our 10- and 11-year-old children, the belief-desire distinction was demonstrated in channel CP6—the same channel that demonstrated the effect in 7- and 8-year-olds—as well as in channel P6—the same channel that demonstrated the effect in adults.

As noted in the introduction, several fMRI studies implicate the right posterior region of the right TPJ as recruited for theory-of-mind reasoning and belief-reasoning in both children (e.g., Sommer et al., 2010; Pfeifer et al., 2009; Saxe et al., 2009; Gweon et al., in press) and adults (see e.g., Carrington & Bailey, 2009 for review). Moreover, some of these studies suggest that the right TPJ becomes increasingly specialized for mental-state reasoning as children age and as accuracy for mental-state reasoning improves (Saxe et al., 2009; Gweon et al., in press).

Our results go further to demonstrate a greater activation for belief-reasoning *compared to desire-reasoning* in right posterior scalp regions, suggesting more strongly that right posterior regions of cortex (such as the right TPJ) actually become specialized for *belief*-reasoning specifically, beyond reasoning even about other mental states such as desires (see also Bowman et al., 2012 and Liu et al., 2009a). Such an interpretation is consistent with the fMRI data presented above, and with our behavioral performance accuracy data demonstrating advancements in accuracy for belief-reasoning (beyond domain general advancements and beyond advancements in accuracy for desire-reasoning) as children age from middle to late childhood. As outlined at the outset of the discussion section, such a pattern of neural specialization could suggest a possible neural mechanism by which belief-understanding may build off of desire-understanding, with additional neural substrates in right posterior regions recruited for specifically belief-reasoning as children age and advance in their understanding of complex mental-states (such as beliefs). A critical future direction would thus be to verify that the right posterior regions implicated in the present study and in Bowman et al. and Liu et al. indeed correspond to the right TPJ, in order to clarify how our ERP neural activation patterns fit with the fMRI patterns. ERP source localization methods as well as functional near-infrared spectroscopy (fNIRS) methods which allow for unambiguous and more precise localization techniques may be particularly useful in this endeavor (as demonstrated in Study 2 of this dissertation).

Age-related differences in latency of the right posterior belief-desires distinction. Though on the whole, the right posterior belief-desires distinction was similar across the 7- and 8-year-old (Bowman et al. 2012), 10- and 11-year-old (present study), and adult (Liu et al., 2009a) samples, there were some intriguing differences in the timing or latency of the effect. Specifically, in the present study and in the younger 7- and 8-

year-old sample (Bowman et al.), the belief-desire distinction occurred at the tail end of the ERP time course, up to 1400 ms. Whereas for adults (Liu et al.), the distinction was demonstrated earlier in the time course, in the 600-800 ms epoch. This age-related latency pattern is in line with prior ERP research comparing false-belief reasoning in young children versus adults (Liu, Sabbagh, Gehring, & Wellman, 2009b). Indeed, in general, younger age groups commonly show effects that are similar to older age groups but that occur *later* in the ERP time course (DeBoer et al., 2004; Taylor & Baldeweg, 2002). In particular, this pattern is demonstrated in developmental examinations of memory (e.g., Czernochowski, Mecklinger, Johansson, & Brinkmann, 2005), and thus it has been argued that age-related latency differences likely reflect the development of faster information processing (Kail, 1991).

However, there was an additional timing difference between the present study's 10- and 11-year-old sample and the 7- and 8-year-old sample from Bowman et al. (2012) that does not fit a pattern of age-related decreases in latency of effects. Specifically, though in these younger 7- and 8-year-old children, the belief-desire distinction occurred late in the ERP time course at the tail end of the waveform, the effect was also demonstrated earlier in the 600-800 ms epoch. An interesting question thus concerns why the present study's 10- and 11-year-old sample would *not* show a belief-desire distinction in the 600-800 ms epoch, when it was found in this epoch in both the 7- and 8-year-old (Bowman et al.) and adult (Liu et al., 2009a) samples.

One possibility is that the pattern of effects across the three age groups represents a developmental shift from implicit (non-conscious) belief-reasoning earlier in childhood, to slow, conscious and deliberate reasoning later in childhood—reasoning which then gets faster and more efficient in adulthood. Both behavioral research and the neural ERP data provide some support for this possibility.

Several behavioral studies now demonstrate that before developing the capacity for verbal, explicit reasoning, early on in the second year of life babies reveal some implicit, non-verbal capacity for processing scenes depicting characters' beliefs and false-beliefs (e.g., Onishi & Baillargeon, 2005; Southgate, Senju & Csibra, 2007). Moreover, by 2 and 3 years of age children are able to solve explicit (verbal) mental-state reasoning tasks that require simple judgments about actions based on a character's desire states, but have difficulty using similar explicit reasoning about characters' belief states (e.g., Wellman & Liu, 2004). There is considerably less research on behavioral theory-of-mind reasoning in adults, however some research suggests that for adults, belief-reasoning comes to represent a more automatic process evidenced by faster response times on false-belief tasks versus tasks requiring reasoning about non-mental representations that can also be false (e.g., signs, maps) (Cohen & German, 2011; see also German & Cohen, 2011 for review). Thus, the behavioral data suggest a progression of belief-reasoning that develops from passive and implicit, to complex and deliberate, to faster and automatic processing.

A comparison of the ERP data across the 7- and 8-year-old (Bowman et al., 2012), 10- and 11-year-old (present study) and adult (Liu et al., 2009a) samples provides some intriguing indirect support for the possibility that earlier in development, children rely on more passive, implicit cognitive processes to reason about beliefs but use more complex and deliberate cognitive reasoning later in development. Specifically, for the 7- and 8-year-old sample, while the belief-desire-distinction first reached full significance in the 600-800 epoch, the effect reached near significance as early as 350 ms post stimulus (Bowman et al., 2012). In contrast, for adults, though the belief-desire distinction began in the 600 ms epoch, it was most robust later in the time course in the 700-800 ms epoch (Liu et al., 2009a). This pattern of results therefore demonstrates that, in the 7- and 8-year-old sample, belief-reasoning emerged as distinct from desire-reasoning considerably earlier in the ERP time course compared to adults, and provides some evidence that the belief-reasoning process for these youngest children—captured by this early distinction—is potentially substantively different from the reasoning process captured by a later distinction in the adult (and older child) age groups.

Indeed, critically, researchers argue that *earlier* ERP components represent more passive, less-involved cognitive processing, whereas *later* components involve more complex cognitive reasoning and evaluation (e.g., Davies, Chang, Gavin, 2010; Polich, 1993). Thus, the belief-desires distinction that began as early as 350 ms and extended through to the 600-800 ms epoch in the 7- and 8-year-old sample (Bowman et al., 2012) could very well reflect an 'early' ERP component of passive, implicit cognitive processing. Likewise, the effect that occurred at the tail end of the ERP waveform in the 1300-1400 ms epoch in both the 7- and 8-year-old sample and in the present study 10- and 11-year-old sample would reflect a 'later' ERP component involving more complex cognitive reasoning and evaluation. This same complex-reasoning component could also be reflected in the adult sample (Liu et al., 2009a) by the effect that began 600 ms post stimulus and became most robust in the 700-800 ms epoch. That is, for adults, an effect in the 600-800 ms epoch likely represents a 'late' effect (and indeed was classified as such in Liu et al.). This 'late' effect in adults is just earlier than the 'late' effect in the child samples due to a general increase in speed of processing across development, as outlined above (e.g., Czernochowski, et al., 2005; Kail et al., 1991).

Suppose, speculatively, that the 10- and 11-year-olds in the present study engaged less in ‘passive’, ‘implicit’ processing (evidenced by a *lack* of belief-desire distinction early in the ERP time course), and relied more heavily on more involved reasoning and evaluation (that occurred later in the ERP time course) to solve the belief-reasoning tasks. If so, a heavier reliance on a slower deliberate type of reasoning and less use of quicker, passive, implicit processing could explain, in part, why these older children showed much improvement in accuracy for belief-reasoning compared to the younger child sample (who may have relied more heavily on a quicker, passive, and therefore potentially less accurate system of cognitive processing, evidenced by a belief-desire distinction early in the ERP waveform).

In order to more directly test the possibility that different types of belief-reasoning (e.g., passive and implicit versus complex and deliberate) are reflected differently in ERP components, investigations of the neural correlates of belief- and desire-reasoning in preschool aged children seem especially critical. Such investigations would allow insight into the neural correlates of belief- and desire-reasoning during the time in which children are undergoing stark behavioral developments in both belief- and desire-reasoning—developments that include advancements from capacities for simple judgments to capacities for more complex explanative reasoning. Thus, investigations of these younger children would help shed light on the role of accuracy, and on the role of implicit versus explicit/explanative reasoning, in the neural processes supporting the development of theory of mind.

Conclusions

The present study investigated the neural correlates of belief- and desire-reasoning in 10- and 11-year-old children using ERP methods. Our data provided a developmental comparison to two prior parallel ERP studies examining belief- and desire-reasoning in 7- and 8-year-old children (Bowman et al., 2012) and adults (Liu et al., 2009a). Results of the present study revealed patterns of belief and desire activations that were comparable to both the younger and older age groups: similar to younger children and adults, 10- and 11-year-olds demonstrated both a common system for reasoning about mental-states more generally (evidenced by both belief and desire activations differentiating from the physical control, but not differentiating from each other), as well as an additional pattern of activation for belief-reasoning—distinct from activation for desire-reasoning—that could suggest development of a separate system for reasoning about specifically beliefs.

Our 10- and 11-year-old data also revealed several intriguing differences from both the younger and older age groups in terms of location, timing, and polarity of effects. These differences suggest that the 10- and 11-year-old sample may be capturing a transitional state in the neural correlates of belief- and desire-reasoning as they develop from middle childhood to adulthood. Moreover, the differences reveal that the mechanisms underlying theory-of-mind development are complex, and lead to questions of potential reorganization of underlying neural substrates, developments in speed of cognitive processing, and changes in types of cognitive processing used at different stages of development that are all ripe for future research.

Table 5.1

Means, Standard Deviations, and T-tests Comparing Beliefs, Desires and Physical Condition Accuracy across the Younger (7- & 8-year-old; Bowman et al., 2012) and Older (10- & 11-year-old) Age Groups

Condition Accuracy	7&8 Years Mean (SD)	10&11 Years Mean (SD)	Age Group *t*-test
Beliefs	.65 (.12)	.81 (.15)	$t(34) = -3.46, p = .002^{**}$
Desires	.86 (.06)	.92 (.06)	$t(34) = -2.70, p = .010^{**}$
Physical	.88 (.08)	.91 (.07)	$t(34) = -1.26, p = .216$

Notes. $^{**}p < .01$, SD = standard deviation, sample sizes for 7&8 years and 10&11 years sample are N=16 and N=20, respectively.

Table 5.2

Correlations between Age, Beliefs Accuracy, Desires Accuracy, and Physical Accuracy Across both 7- & 8-year-old (Bowman et al. 2012) and 10- & 11-year-old Age groups

Variable	Beliefs Accuracy	Desires Accuracy	Physical Accuracy
Age	.47**	.44*	.36*
Beliefs Accuracy		.40*	.26
Desires Accuracy			.61**

Notes. ** $p < .01$, *$p < .05$

Table 5.3

Regression Coefficients for Models Predicting Beliefs, Desires, and Physical Accuracy

Dependent Variable	Significant Predictor			Excluded Variables		
	Variable (Std. Beta)	*t*-value	*p*-value	Variable (Beta in)	*t*-value	*p*-value
Beliefs Accuracy	Age (.47)	3.10	.004**	Desires Accuracy (.25)	1.48	.249
				Physical Accuracy (.11)	.66	.515
Desires Accuracy	Physical Accuracy (.51)	4.52	.000***	Beliefs Accuracy (.26)	1.94	.060
				Age (.26)	1.82	.078
Physical Accuracy	Desires Accuracy (.74)	4.52	.000***	Age (.10)	.67	.510
				Beliefs Accuracy (.01)	.09	.931

Note. *** $p < .001$, **$p < .01$, *$p < .05$, Std. Beta = standardized Beta

Table 5.4

Frontal ROI Cluster Analysis Results for the 2(condition) x 5(channel) Repeated Measures ANOVAs for 4 Epochs of Interest

	Epoch (ms post stimulus onset)			
Comparison	200-250	350-500	600-800	800-850
	Mid-Frontal Cluster			
Belief vs. Desire	*F*(1,14) = .352, *p* = .562	*F*(1,14) = .196, *p* = .668	*F*(1,14) = 1.82, *p* = .199	*F*(1,14) = .16, *p* = .694
Belief vs. Physical	*F*(1,14) = 14.59, *p* = .002**	*F*(1,14) = 12.98, *p* = .003**	*F*(1,14) = 2.60, *p* = .130	*F*(1,14) = 2.31, *p* = .151
Desire vs. Physical	*F*(1,14) = 32.17, *p* < .001***	*F*(1,14) =9.41, *p* = .008**	*F*(1,14) = 7.25, *p* = .017*	*F*(1,14) = 3.37, *p* = .080†
	Right Mid-Frontal Cluster			
Belief vs. Desire	*F*(1,14) = 1.05, *p* = .323	*F*(1,14) = .017, *p* = .899	*F*(1,14) = .40, *p* = .538	*F*(1,14)= .004, *p* = .949
Belief vs. Physical	*F*(1,14) = 15.80, *p* = .001**	*F*(1,14) = 4.57, *p* = .051†	*F*(1,14) = 1.36, *p* = .262	*F*(1,14) = 1.58, *p* = .230
Desire vs. Physical	*F*(1,14) = 45.54, *p* < .001***	*F*(1,14) =5.94, *p* = .029*	*F*(1,14) = 3.37, *p* = .088†	*F*(1,14) = 1.61, *p* = .226

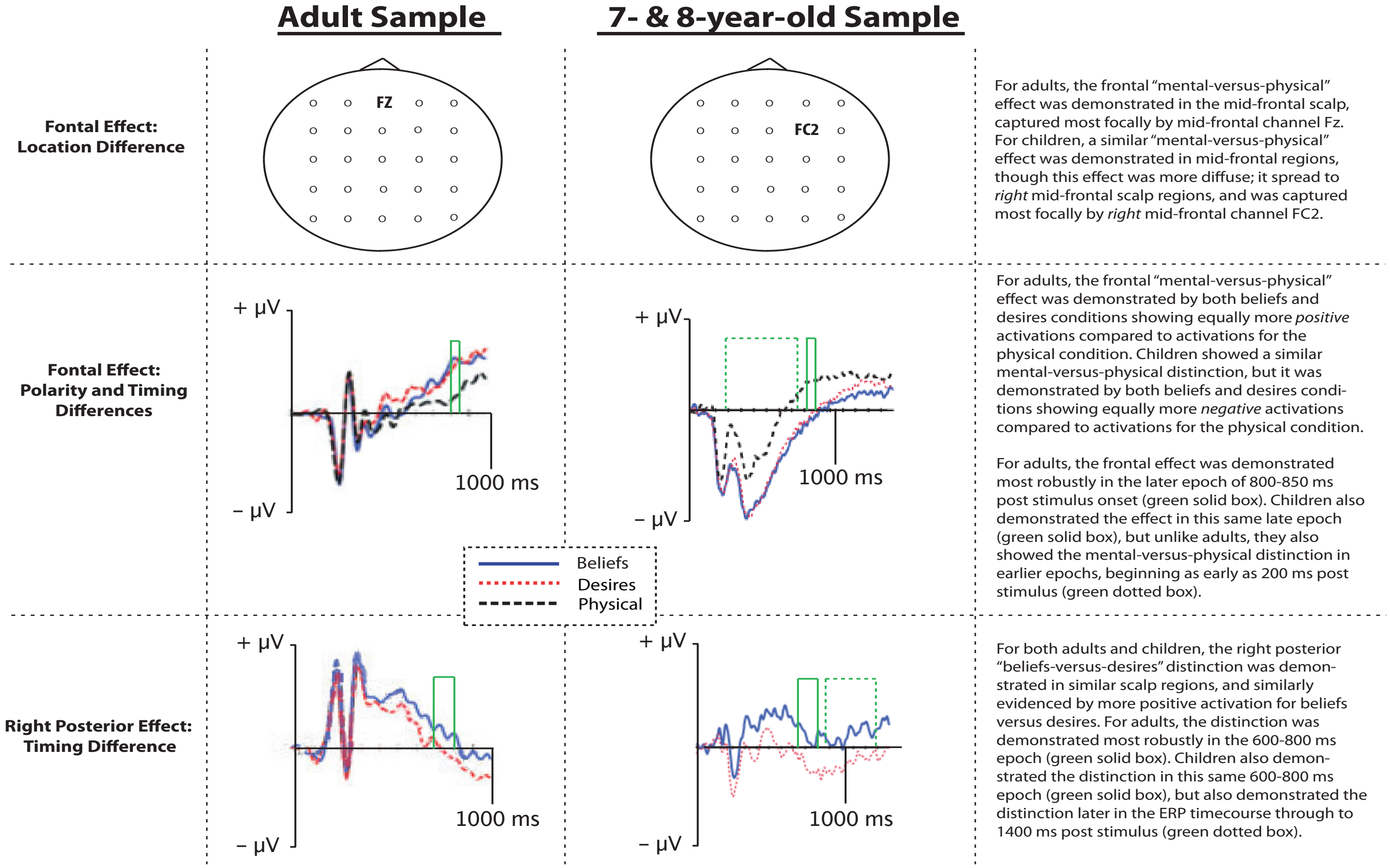
Notes. *** *p* < .001, ** *p* < .01, * *p* < .05, †, *p* < .1.

Table 5.5

Right Posterior ROI Cluster Analysis Results for the 2(condition) x 4(channel) Repeated Measures ANOVAs for 3 Epochs of Interest

	Epoch (ms post stimulus onset)		
Comparison	600-800	1100-1200	1300-1400
Belief vs. Desire	$F(1,14) = .15, p = .709$	$F(1,14) = 1.30, p = .273$	$F(1,14) = 5.76, p = .031$*
Belief vs. Physical	$F(1,14) = .01, p = .917$	$F(1,14) = .07, p = .792$	$F(1,14) = 1.32, p = .270$
Desire vs. Physical	$F(1,14) = .06, p = .812$	$F(1,14) = 1.79, p = .202$	$F(1,14) = .54, p = .475$

Note. * $p < .05$



For adults, the frontal "mental-versus-physical" effect was demonstrated in the mid-frontal scalp, captured most focally by mid-frontal channel Fz. For children, a similar "mental-versus-physical" effect was demonstrated in mid-frontal regions, though this effect was more diffuse; it spread to *right* mid-frontal scalp regions, and was captured most focally by *right* mid-frontal channel FC2.

For adults, the frontal "mental-versus-physical" effect was demonstrated by both beliefs and desires conditions showing equally more *positive* activations compared to activations for the physical condition. Children showed a similar mental-versus-physical distinction, but it was demonstrated by both beliefs and desires conditions showing equally more *negative* activations compared to activations for the physical condition.

For adults, the frontal effect was demonstrated most robustly in the later epoch of 800-850 ms post stimulus onset (green solid box). Children also demonstrated the effect in this same late epoch (green solid box), but unlike adults, they also showed the mental-versus-physical distinction in earlier epochs, beginning as early as 200 ms post stimulus (green dotted box).

For both adults and children, the right posterior "beliefs-versus-desires" distinction was demonstrated in similar scalp regions, and similarly evidenced by more positive activation for beliefs versus desires. For adults, the distinction was demonstrated most robustly in the 600-800 ms epoch (green solid box). Children also demonstrated the distinction in this same 600-800 ms epoch (green solid box), but also demonstrated the distinction later in the ERP timecourse through to 1400 ms post stimulus (green dotted box).

Figure 5.1. Summary of ERP component differences for adult (Liu et al., 2009a) and 7- and 8-year-old (Bowman et al., 2012) samples.

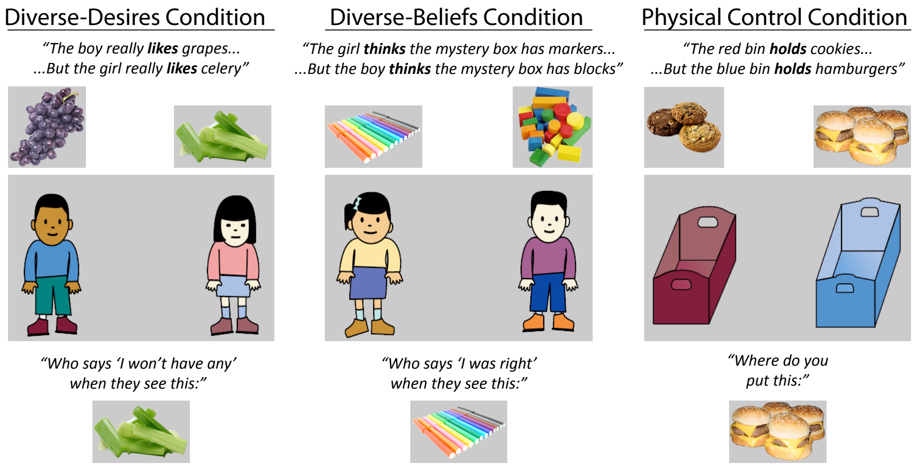


Figure 5.2 (Taken directly from Bowman et al., 2012). A schematic of the theory of mind Diverse-Desires (left), Diverse-Beliefs (middle), and Diverse-Physical (right) tasks showing examples of information phase (top) and target questions (bottom) as well as sample graphics for both food and toy condition types.

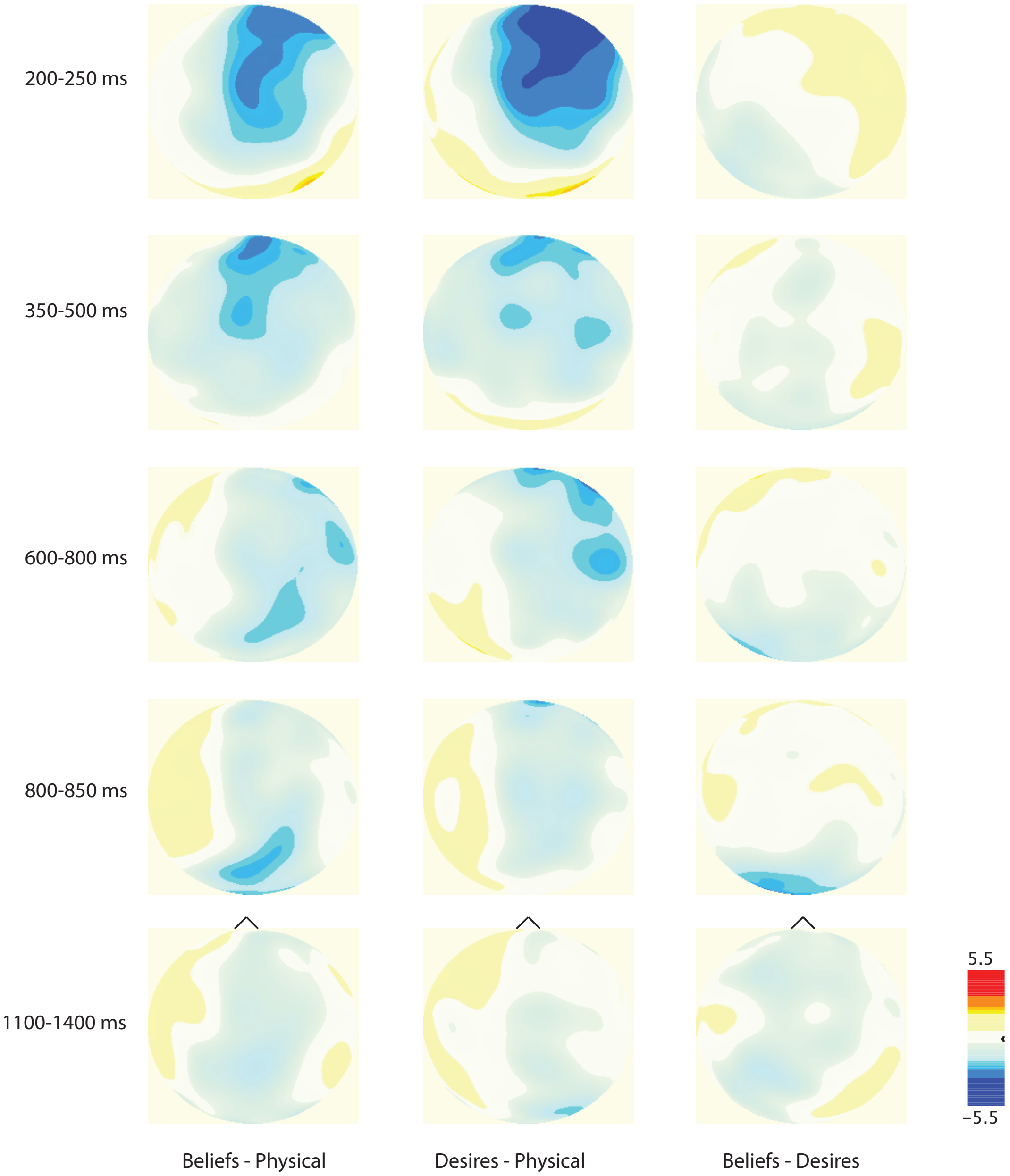


Figure 5.3. Topographical maps of the scalp electrical activity at all 128 channels. Maps depict mean amplitude difference for Physical subtracted from Beliefs (left), Physical subtracted from Desires (middle), and Desires subtracted from Beliefs (right) in the 200-250 ms (top), the 350-500 ms, 600-800 ms, 800-850 ms, and 1100-1400 ms (bottom) post stimulus epochs. Maps are oriented with frontal position up from an overhead scalp view. Darker regions indicate greater negative differences between two conditions. The maps show a clear difference in the 200-250 and 350-500 ms epochs for the Beliefs-Physical and Desires-Physical comparisons (Beliefs and Desires more negative compared to Physical) in the frontal channels, with this difference diminishing in later epochs.

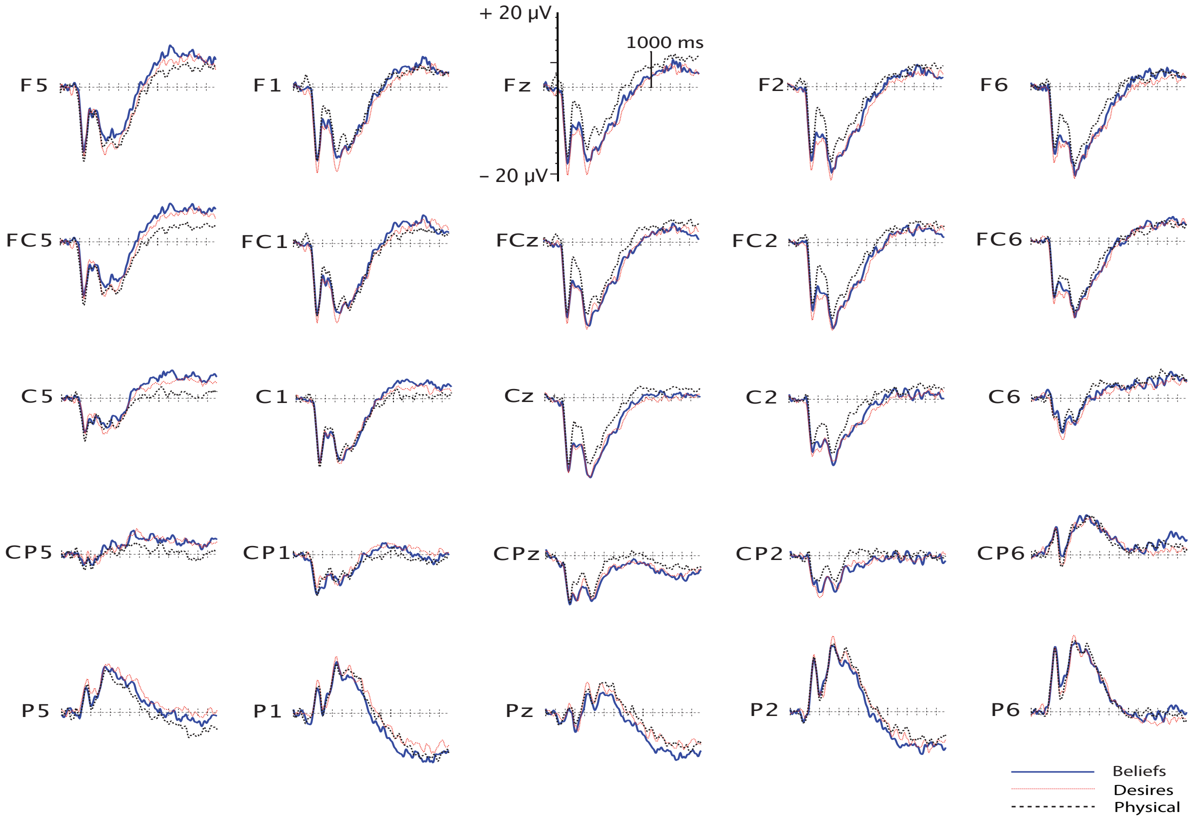


Figure 5.4. Grand average event-related brain potential waveforms for the Beliefs (blue solid lines), Desires (red dotted lines), and Physical (black dashed lines) conditions from 25 channels selected to encompass locations from left to right and front to back (top to bottom in the grid) over the scalp. The 128 EGI channels can be grouped and labeled in terms of the 10-10 system convention. This figure adopts that more standard convention, with the selected channels reflecting 10-10 labels and positions. Waveforms are displayed with positive amplitudes above the axis and negative amplitudes below.

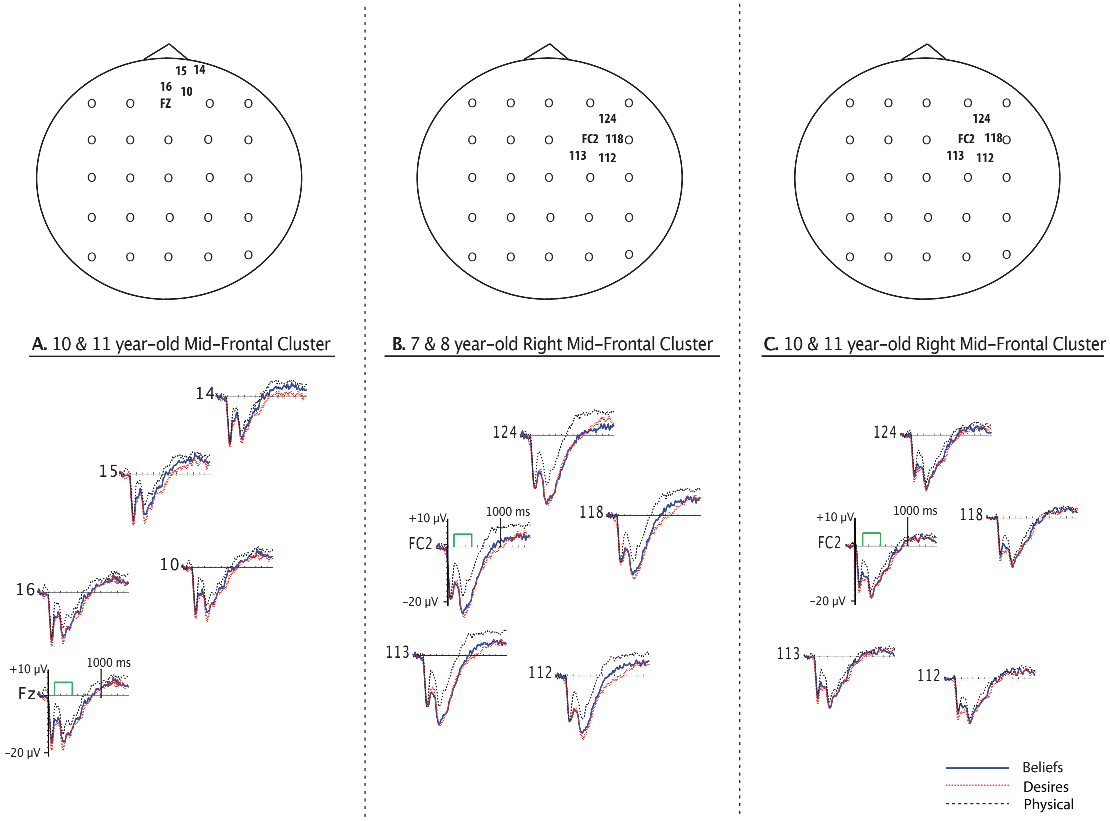


Figure 5.5. Grand average event-related brain potential waveforms for the Beliefs (blue solid lines), Desires (red dotted lines), and Physical (black dashed lines) for the frontal 5-channel cluster. Panel A (left) depicts the *mid-frontal* cluster for the 10- and 11-year-old (present study) sample. Panel B (middle) depicts the *right-mid-frontal* cluster for the 7- and 8-year-old sample (Bowman et al., 2012). Panel C (right) depicts this same *right-mid-frontal* cluster but for the 10- and 11-year-old (present study) sample. The solid green boxed section indicates epochs in which the present study sample showed the strongest significant 'mental-versus-physical' effect with Beliefs and Desires conditions equally differentiating from the Physical condition. Waveforms are displayed with positive amplitudes above the axis and negative amplitudes below. Scalp positions for the 5-channel cluster (bold channels) relative to the 5 x 5 grid are shown at the top.

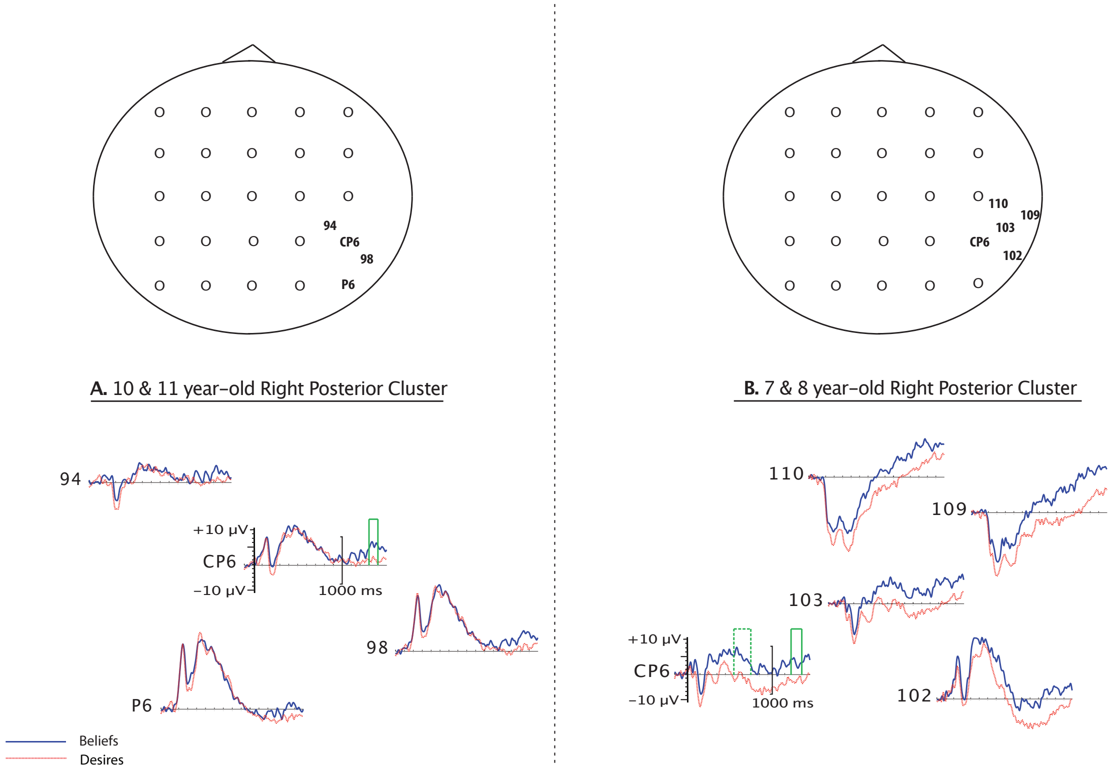


Figure 5.6. Grand average event-related brain potential waveforms for the Beliefs (blue solid lines) and Desires (red dotted lines) for the right posterior channel clusters. Panel A (left) depicts the right posterior 4-channel cluster for the 10- and 11-year-old (present study) sample. Panel B (right) depicts the right posterior 5-channel cluster for the 7- and 8-year-old sample (Bowman et al., 2012). The solid green boxed section indicates epochs in which the present study sample showed the strongest 'belief-versus-desire' effect with amplitude for belief activation significantly greater than desire activations. The green dotted box indicates the earlier epoch in which the 7- and 8-year-old sample also showed a similar beliefs-desires distinction. Waveforms are displayed with positive amplitudes above the axis and negative amplitudes below. Scalp positions for the channel clusters relative to the 5 x 5 grid are shown at the top. Notably, both younger and older sample clusters surround channel CP6, demonstrating a similar location of effect.

CHAPTER VI

General Discussion

The three studies in this dissertation seek to capture a more comprehensive picture of the process of theory-of-mind development by examining the neural correlates of theory-of-mind reasoning in child populations. Taken together, theses studies examine theory-of-mind neural correlates over early, middle, and late childhood, substantially increasing the currently small pool of pediatric neurocognitive studies. At the broadest level, they highlight the utility of a neuroscientific approach to investigations of theory of mind, and illustrate the importance of using this approach to collect *developmental* data that spans multiple age ranges, and in particular, that spans ages encompassing childhood development.

Utility of the Neuroscientific Approach

As outlined in the general introduction, behavioral research frames our current understanding of changes and advances in theory-of-mind reasoning, yet, there are limits to what behavioral research alone can teach us. The three studies of this dissertation demonstrate how cognitive neuroscientific methods can shed light on the mechanisms and processes that underlie, influence, and pace development, which can in turn help address developmental questions left unanswered by the behavioral data alone. In particular, the studies in this dissertation shed light on the processes underlying the progression from an early understanding of desires to a later understanding of beliefs, demonstrated clearly in the behavioral research in early childhood (e.g., Wellman & Liu, 2004). And most broadly, they shed light on continuity and change in theory-of-mind development—one of the most fundamental and yet under-explored issues in the theory-of-mind development literature.

Shedding Light on the Progression from Understanding Desires to Understanding Beliefs

As noted in the general introduction and in the reviews of current research throughout the chapters, existing neuroscientific investigations have focused primarily on belief- and false-belief reasoning or on undifferentiated mental-state reasoning more generally. However, these foci are limited; theory of mind—often termed a belief-desire or belief-desire-emotion naïve psychology—involves understanding multiple, causally interconnected mental concepts, as evidenced most strongly in the behavioral data by the robust progression from understanding desires to understanding beliefs demonstrated over the preschool years (e.g., Wellman & Liu, 2004). Studies 2 and 3 include a comparison of the neural correlates of belief- *and* desire-reasoning in the same sample of children, directly measuring similarities and differences in underlying developmental processes supporting reasoning about these two mental states.

Critically, both studies demonstrated that at ages 7 and 8 years (Study 2, fNIRS), and 10 and 11 years (Study 3, ERP), the neural activations for reasoning about beliefs diverge from those for reasoning about desires in right TPJ and right posterior scalp regions. This distinction between the neural mechanisms underlying these two types of reasoning stands in contrast to the developmental picture given by the standard behavioral measures of belief- and desire-reasoning, which yield ceiling performance on both belief and desire tasks after age 6 years (e.g., Wellman & Liu, 2004). Behavioral performance on the parallel fNIRS and ERP tasks employed in these two studies did reveal differences in accuracy for belief- and desire-reasoning at these older ages. Thus, one might argue that the use of more sensitive behavioral measures beyond early childhood would lead to similar conclusions about differential development of belief- and desire-reasoning even into middle and late childhood. However, the neural data from Study 3 reveals that such a conclusion would still be incomplete and inaccurate, and that in fact the neural processes underlying belief- and desire-reasoning are not that straightforward: Though accuracy for belief-reasoning was lower compared to accuracy for desire-reasoning at 10 and 11-years-old, both belief- and desire-reasoning yielded an additional pattern of activation in the mid-frontal scalp, in which the activations for the two types of reasoning were comparable and indeed overlapping—suggesting the existence of a neural system common to both types of reasoning, despite behavioral performance differences.

Thus, the neural data from Studies 2 and 3 demonstrate a crucial advantage of a neuroscientific approach to studying development: neural data can reveal differences in cognitive processes where behavioral data suggest similarities, and they can reveal similarities in cognitive processes despite behavioral performance differences. Most valuable to the study of theory-of-mind development, these identified differences and similarities in neural activation patterns for belief- and desire-reasoning shed light on the currently outstanding question of what might underlie the behavioral progression from understanding desires to understanding beliefs. Specifically, the pattern of *both* a) a common neural system for reasoning about beliefs and desires in mid frontal regions, and b) a right posterior system for reasoning about beliefs as distinct from desires represents a straightforward developmental possibility that an understanding of beliefs may build on prior desire-understanding, evidenced by additional substrates in right TPJ regions being recruited for specifically belief reasoning (beyond recruitment for desire-reasoning) as belief-understanding becomes more distinct and accurate. Indeed, Study 2 clarified that, at least for 7- and 8-year-olds, this right posterior beliefs-desires distinction occurs in the right TPJ, and represents truly *greater* activation for belief- compared to desire-reasoning. Study 2 thus helps to situate these belief-desire activation patterns in the current fMRI investigations of theory-of-mind reasoning in children, which suggest an increased specialization of the right TPJ for mental-state processing as children age from middle to late childhood and as accuracy for theory-of-mind reasoning improves (Saxe et al., 2009; Gweon et al., in press). Therefore, the neural results from Study 2, in conjunction with this existing fMRI research, provide additional support for the developmental possibility outlined above. Taken together, results from Studies 2 and 3 suggest a neural mechanism underlying the progression from an early desire-understanding to a later understanding of beliefs.

Shedding Light on Continuity and Change in Theory-of-Mind Development

The advantage of a more direct and clear measure of underlying similarities and differences in cognitive processing, conferred by a neuroscientific approach, also serves to shed light on elements of continuity and change in development. Indeed, a comparison of the neural correlates of theory-of-mind reasoning across development highlights similarities and differences in underlying neural mechanisms that can reveal continuity and change in cognitive reasoning, and that can point to factors that may govern these developmental patterns.

Study 1 most clearly illustrates the utility of neuroscientific data in highlighting patterns of continuity and change in theory-of-mind development. Study 1 tracked the neural correlates of theory-of-mind reasoning longitudinally over early to middle childhood, and demonstrated two sorts of continuity in the neural substrates associated with theory of mind: 1) children's behavioral theory-of-mind performance measured at age 4 years positively predicted the extent to which their DMPFC exhibited selectivity for mental-state reasoning at age 7 and 8 years, and 2) functional maturation of regions of the DMPFC measured at 4 years also positively predicted the extent to which those same regions of the DMPFC were selective for mental-state reasoning at age 7 and 8 years. Knowledge of the existence of such a clear effect of continuity informs future research to investigate the developmental factors that contribute to such an effect. For example, the robustness and spatial precision of the effect could suggest that one such factor may be biological or genetic in nature. Indeed, given the documented role of dopamine in the development and maintenance of cells in the MPFC (Kalsbeek et al., 1987; Popolo et al., 2004) and the relation between dopamine and 4-year-olds' behavioral theory-of-mind performance (Lackner et al., 2010; 2012), the continuity in the DMPFC outlined in Study 1 suggests that dopamine may play a role in the development and maintenance of the relation between DMPFC and theory of mind over early to middle childhood, and thus future longitudinal research on theory-of-mind neural correlates should include examinations of potential genetic or dopaminergic contributions.

Though Study 3 does not provide longitudinal data, it does provide an important developmental comparison to parallel investigations of belief and desire neural correlates in younger (7- and 8-year-old, Bowman et al., 2012) and older (adult, Liu et al., 2009a) samples. The comparisons of belief and desire neural activation patterns across these three age groups reveal, at least indirectly, elements of both continuity and change in the neural mechanisms underlying belief- and desire-reasoning. Most broadly, all three age groups demonstrated overlapping activations for belief- and desire-reasoning in mid-frontal scalp regions, *and* distinct activations for belief- versus desire-reasoning in right posterior scalp regions. These results clearly indicate an important continuity in the neural mechanisms underlying belief and desire-reasoning from middle childhood, to late childhood, and potentially even to adulthood: for all three age groups, two neural systems for reasoning about beliefs and desires exist.

These results further raise the need for future research to investigate belief and desire neural correlates in early childhood, when belief- and desire-reasoning demonstrate stark behavioral differences on the standard tests of theory of mind. Such future studies can now investigate whether similar neural systems exist even earlier in development, or whether there is some change in the underlying neural correlates that accompanies the stark behavioral change going on at these earlier stages, as children transition from early to middle childhood. Indeed, results of Study 1 suggest continuity in mid-frontal neural substrates (i.e., the DMPFC), but reveal a distinct *lack* of continuity in right posterior neural substrates (i.e., right TPJ), pointing to the possibility that the right posterior belief-desire distinction demonstrated in middle childhood and beyond might be different in early childhood. Thus, the studies in this dissertation demonstrate how both tracking neural systems across development longitudinally, and comparing neural correlates of equivalent cognitive tasks across different discrete age groups can reveal patterns of continuity and change—patterns that serve to illuminate the process by which cognitive understanding, and specifically a theory of mind, develops.

Importance of Developmental Data

Much can be revealed from investigations of the neural correlates of cognition, as outlined in the previous section. Yet, also as outlined above, neurocognitive investigations are even more valuable when they include examinations at multiple different stages of development. Indeed, the patterns of continuity and change highlighted above were only revealed when neural data was compared across multiple developmental time points. Likewise, though we have learned much about the neural mechanisms that might underlie a progression from desire-understanding to an understanding of beliefs from three points of cross-sectional data, further data is needed from even younger children to better understand the processes governing this developmental progression as it is occurring in early childhood.

Indeed, critically, the studies in this dissertation demonstrate that it is only with developmental data that a comprehensive picture of development, in all of its complexities, can be revealed. For example, Study 1 provides a follow-up examination of the theory-of-mind neural correlates initially assessed when children were 4-years-old. That initial assessment (by Sabbagh et al., 2009) showed that children's behavioral theory-of-mind performance positively related to functional maturation of both the DMPFC and RTPJ—two neural regions that are consistently implicated in neuroscientific investigations of theory of mind in adults (e.g., Carrington & Bailey, 2009; Apperly, 2011). Thus, on the basis of this most broad child-adult comparison, one might conclude that there is stark continuity in each of these neural regions as related to theory of mind from early childhood through adulthood.

However, in Study 1, the follow-up examination of those children three years later at 7- and 8-years-old demonstrated direct evidence for continuity in the relation between DMPFC and theory of mind, but also indirect evidence for *change* in the relation between theory of mind and the RTPJ. Thus, this additional piece of developmental data reveals a more complex developmental process governing theory of mind, at least as related to the RTPJ. Indeed, other developmental investigations examining theory of mind neural correlates in childhood (e.g., Saxe et al., 2009; Gweon et al., in press), including Studies 2 and 3, illustrate similar complexities with respect to how the RTPJ supports theory-of-mind reasoning. Thus, as different stages of development are more closely examined, the further complexities of the developmental process are revealed.

Results from Study 3 also clearly illustrate the importance of developmental data in uncovering a fully complex developmental process. The examination of belief and desire neural correlates in 10- and 11-year-old children provided an additional window on the development of the neural mechanisms underlying belief- and desire-reasoning, offering clarification on how these correlates develop between middle childhood (i.e., 7- and 8-years-old, Bowman et al., 2012) and adulthood (adults, Liu et al., 2009a). Critically, this additional window revealed that the process of development is complex and highlighted how, even over just a few years of development, the neural mechanisms for belief- and desire-reasoning can undergo much change. Indeed, Study 3 showed that the neural mechanisms supporting belief- and desire-reasoning across age groups change in location, timing, and polarity of neural activations. These changes point to a reorganization of underlying neural substrates, developments in speed of cognitive processing, and potentially even changes in the *type* of cognitive processing (e.g., passive/implicit; explicit, slow and deliberate; automatic/efficient) that might be privileged at different stages of development. Thus, the added richness of developmental data provided by Study 3 shed more light on the richness of the underlying developmental process, and served to raise even more questions about how an understanding of different mental states build to form an accurate, and expert theory of mind—questions that future developmental neuroscientific research can begin to address.

Future Directions

Much is now known about the neural correlates of theory of mind. Neuroscientific investigations with adults along with the handful of studies with children demonstrate useful methods, testable hypotheses, and intriguing results. Indeed, the three studies of this dissertation add considerably to the handful of studies with children to shed valuable light on continuity and change in the neural regions supporting theory-of-mind development, and on the neural processes underlying how an understanding of beliefs and desires may build to form a complete theory of mind. Yet, as also illuminated by the studies in this dissertation, much is still unknown.

Examinations of Multiple Different Mental States

Further research is needed moving beyond examinations of undifferentiated mental states and a narrowed focus on beliefs, to including direct examinations of desires. Studies 2 and 3 provide some of these initial investigations by including examinations of direct contrasts between neural correlates for belief- versus desire-reasoning. Even more comprehensive examinations would include direct contrasts of other mental states as well, such as emotions and intentions. Each of these mental concepts is interconnected (we can be sad if we don't get what we want; we can be happy if we think someone intended to do something nice), but they also show distinct developments behaviorally (e.g., Wellman & Liu, 2004). Understanding of the neural correlates of beliefs, desires, emotions, and intentions considered together will help illuminate how these mental concepts build and relate to form a complex and fully-fledged theory-of-mind.

Investigations of Theory-of-Mind Neural Correlates in Infancy and Early Childhood

Behaviorally, mental-state understanding develops considerably over the first two years of life. There is clear development of intention-understanding, understanding of knowledge and ignorance, and evidence for some early implicit understanding of beliefs. But the nature of these early achievements as well as the larger developmental picture is highly controversial and unclear. Infant cognitive developments are measured most consistently with non-verbal, implicit, looking-time methods, and disputes rest highly on questions of how these implicit measures relate to the explicit, verbal measures used with older children and adults. Just as existing neuroscientific investigations shed light on differences and similarities between theory of mind and executive functioning processes, belief- versus desire-reasoning, and cognitive processing in adults versus children, so too could they help to distinguish implicit versus explicit types of cognitive processing. Indeed, results of Study 3 suggest that different types of processing might be engaged more heavily at different stages of development (even in middle and late childhood), and raise still further questions about whether or not these different types of processing might be related to advancements in accuracy for mental-state reasoning.

These are important issues within cognitive science and cognitive development more generally. Studies that examine neural correlates as infants perform implicit theory-of-mind tasks, as well as studies that examine neural correlates in early childhood to capture cognitive processing as children undergo stark behavioral developments, could identify neural similarities and differences between known explicit signatures and any implicit neural signatures that emerge.

Methods

Existing research already demonstrates useful methods for examining neurocognitive correlates developmentally. FMRI methods have been instrumental in identifying neural correlates in adults, adolescents, and even children as young as 5-years-old. However, the lengthy experiments, noisy testing environments, high cost of fMRI scans, and the need for virtually motionless scans, make these methods less feasible with young children and infants. Electrophysiological (EEG and ERP) methods—such as those used in wave 1 of Study 1 and in Study 3 of the dissertation—offer a more child-friendly approach. These methods are equally non-invasive, but make no noise, and allow parent and child to remain in close proximity (e.g., infants can sit on the mother's lap).

As for experiment duration, Sabbagh et al. (2009) (wave 1 of Study 1) demonstrates that useful task-independent EEG data (baseline/resting data) can be collected from preschool-age children in just 6 minutes. Moreover, source localization of electrophysiological data can help identify brain regions with increased spatial resolution, when fMRI methods are less feasible. Critically, Study 1 provides some validation for these methods. Specifically, Study 1 demonstrated that the regions of the DMPFC that related to theory-of-mind behavioral performance, spotlighted at 4 years (from source localized EEG alpha coherence; Sabbagh et al., 2009) were almost precisely the same regions that were also recruited for theory of mind at 7 years (as measured by fMRI). Moreover, early functional maturation of these DMPFC regions (measured via task-independent 'resting' EEG methods) longitudinally predicted later selectivity in those same regions (measured via task-dependent fMRI).

Thus, as outlined in Chapter II, not only do the longitudinal findings from Study 1 help cross-validate the methods used at both waves, they also demonstrate that task-*independent* 'maturational' aspects of given neural regions can relate to how those regions are actively recruited for cognitive reasoning as measured by task-*dependent* methods, even across three years of development, within a spatial precision on the order of millimeters. These findings thus endorse the use of source localized resting EEG methods more broadly. Such methods could provide a valuable tool for assessing neural correlates of cognition in populations in which longer, more involved methods (e.g., fMRI and ERP) are less feasible, such as with younger children, infants, and populations with more sever developmental cognitive impairments.

As demonstrated in Study 2, functional near-infrared spectroscopy (fNIRS) methods provide another child-friendly alternative to fMRI. As outlined in the introduction to Study 2, these methods use near-infrared light to measure changes in oxy- and deoxy-hemoglobin concentration, and thus measure neurocognitive correlates via changes in the hemodynamic response function, similar to fMRI. They have the same child-friendly properties as those described for EEG/ERP above, and are even more motion-tolerant. Critically, and also as discussed in Study 2, fNIRS data offer clearer measures of individual variation in neural responses in terms of variance in mean oxy-hemoglobin activation across individual participants, as compared to wave-form amplitude in ERP data. This clearer measure of individual variance potentially allows for examination of correlations between performance accuracy and neural activations for cognitive reasoning. Indeed, the ability to investigate how neural correlates for theory-of-mind reasoning change as a function of accuracy seems crucial in light of results from the studies in this dissertation, and in particular the in light of the behavioral findings from Study 3 demonstrating continued difficulty with belief-reasoning even into late childhood.

Investigations of Atypical Development

Studying development that is atypical or delayed can illuminate development more generally. For example, identifying neural similarities and differences in typical versus delayed development could reveal necessary and sufficient component processes supporting cognitive reasoning, as well as different strategies/approaches to the same cognitive tasks. Much is already known about behavioral theory-of-mind development in individuals with autism spectrum disorder (ASD)—a collection of lifelong neuro-developmental disorders characterized by social impairment. Converging evidence shows both delays in theory-of-mind development (by as much as 5 to 10 years or more), and differing sequences of theory-of-mind development (i.e., understanding hidden emotions before understanding false beliefs) in children with ASD compared to typically developing children (e.g., Peterson et al., 2005). Neuroimaging of adults with autism contributed to identification of the theory-of-mind neural network in adults (e.g., Baron-

Cohen et al., 1999). Only recently, however, have there been functional neuroscientific investigations of social cognition in child ASD populations. In their fMRI study, Kaiser and colleagues (2010) found similarities and differences in neural activation for viewing biological motion in children with ASD versus the unaffected siblings of children with ASD and typically developing children. Electrophysiological and fNIRS methods may be particularly useful for investigating neural correlates of theory-of-mind in ASD populations, given the more child-friendly nature of these methods.

Though neuroscientific investigations of theory of mind in ASD populations could help clarify theory-of-mind development in general, such a pursuit may also lead to more confusion. ASD is plagued not only by impairments in theory-of-mind reasoning, but also by other cognitive impairments, including language and executive functioning deficits (Ozonoff, Pennington & Rogers, 2006). Thus, for ASD populations, it is difficult to decipher differences due to theory-of-mind delays, versus those due to the myriad of other cognitive delays associated with ASD. Late-signing deaf children show delays in theory-of-mind development similar to children with ASD (Peterson et al., 2005). Yet deaf children, in contrast, have peripheral, audiological impairments. Moreover, though theory of mind is delayed in late-signing deaf children compared to earlier signers and hearing children, the sequence of theory-of-mind development is the same (Peterson et al. 2005; Wellman, et al. 2011). Thus, this population may offer a better-matched comparison to typically developing children for neuroscientific research. Moreover, given that deaf children develop the same mental concepts in the same sequence as typically developing children, but just two to eight years later, this population allows examination of neural correlates of *early* theory-of-mind developments in older children.

These more physically and mentally mature older children should be able to sustain longer experimental procedures, and yield less movement artifact, thus they should give better quality neuroscientific data overall. To my knowledge, there have been no investigations of theory-of-mind neural correlates in deaf populations.

References

Abraham, A., Rakoczy, H., Werning, M., von Cramon, D. Y., & Schubotz, R. I. (2010). Matching mind to world and vice versa: Functional dissociations between belief and desire mental state processing. *Social Neuroscience, 5*, 1-18.

Adams, R. B., Rule, N. O., Franklin, R. G., Wang, E., Stevenson, M. T., Yoshikawa, S., Nomura, M., et al. (2009). Cross-cultural reading the mind in the eyes: an fMRI investigation. *Journal of cognitive neuroscience*, *22*, 97–108. doi:10.1162/jocn.2009.21187

Apperly, I. A. (2011). *Mindreaders: the cognitive basis of "theory of mind"*. Hove; New York, NY: Psychology Press.

Apperly, I. A., & Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review, 116*, 953-970.

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*, 1691–1703. doi:10.1111/j.1467-8624.2011.01635.x

Aslin, R. N. (2012). Questioning the questions that have been asked about the infant brain using near-infrared spectroscopy. *Cognitive neuropsychology*, *29*, 7–33. doi:10.1080/02643294.2012.654773

Baron-Cohen, S. (2001). Theory of mind in normal development and autism. *Prisme, 34,* 174-183.

Baron-Cohen, S., Ring, H. A., Wheelwright, S., Bullmore, E. T., Brammer, M. J., Simmons, A., & Williams, S. C. (1999). Social intelligence in the normal and autistic brain: An fMRI study. *The European journal of neuroscience*, *11*, 1891-8.

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

Boas, D. A., Dale, A. M., & Franceschini, M. A. (2004). Diffuse optical imaging of brain activation: approaches to optimizing image sensitivity, resolution, and accuracy. *NeuroImage*, *23*, S275–88. doi:10.1016/j.neuroimage.2004.07.011

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,* 618-632. doi:10.1111/j.1467-7687.2012.01158.x

Brainard, D. H. (1997). The psychophysics toolbox, *Spatial Vision 10,* 433-436.

Brandone, A. C., & Wellman, H. M. (2009). You can't always get what you want: Infants understand failed goal-directed actions. *Psychological Science, 20*, 85-91.

Callaghan, T., Rochat, P., Lillard, A., Claux, M. L., Odden, H., Itakura, S., et al. (2005). Synchrony in the onset of mental-state reasoning: Evidence from five cultures. *Psychological Science, 16,* 378-384.

Carlson, S. M., Davis, A., & Leach, J. G. (2005). Less is more: Executive function and symbolic representation in preschool children. *Psychological Science, 16*, 609–616.

Carlson, S. M., & Moses, L. J. (2001) Individual differences in inhibitory control and children's theory of mind. *Child Development, 72,* 1032-1063.

Carrington, S. J., & Bailey, A. J. (2009). Are there theory of mind regions in the brain? A review of the neuroimaging literature. *Human brain mapping*, *30*, 2313-35.

Casey, B. J., Giedd, J. N., & Thomas, K. M. (2000). Structural and functional brain development and its relation to cognitive development. *Biological Psychology. Special Issue: Behavior, Event-Related Potentials, and Neuroimaging: A Lifespan Perspective, 54*, 241-257.

Cohen, A. S., & German, T. C. (2010). A reaction time advantage for calculating beliefs over public representations signals domain specificity for " theory of mind ". *Cognition*, *115*, 417–425. doi:10.1016/j.cognition.2010.03.001

Cooper, R. J., Caffini, M., Dubb, J., Fang, Q., Custo, A., Tsuzuki, D., Fischl, B., et al. (2012). Validating atlas-guided DOT: A comparison of diffuse optical tomography informed by atlas and subject-specific anatomies. *NeuroImage*, *62*, 1999–2006. doi: 10.1016/j.neuroimage.2012.05.031

Czernochowski, D., Mecklinger, A., Johansson, M., & Brinkmann, M. (2005). Age-related differences in familiarity and recollection : ERP evidence from a recognition memory study in children and young adults. *Cognitive, Affective, & Behavioral Neuroscience, 5*, 417–433.

Davies, P. L., Chang, W., Gavin, W. J., & Simon, S. A. (2010). Middle and late latency ERP components discriminate between adults, typical children, and children with sensory processing disorders. *Frontiers in Integrative Neuroscience, 4*, 1–9. doi:10.3389/fnint.2010.00016

DeBoer, T., Scott, L., & Nelson, C. A. (2004). Event-related potentials in developmental populations. In T. Handy (Ed.), *Event-related potentials: A methods handbook* (pp. 263–297). Cambridge, MA: MIT Press.

D'Entremont, B., Hains, S. M. J., & Muir, D. W. (1997). A demonstration of gaze following in 3- to 6-month-olds. *Infant Behavior & Development, 20*, 569-569.

Dunn, L. M., & Dunn, L. M. (1997). *Peabody picture vocabulary test* (3rd ed.). Bloomington, MN: Pearson Assessments.

Ferree, T.C., Clay, M.T., & Tucker, D.M. (2001). The spatial resolution of scalp EEG. *Neurocomputing, 38,* 1209–1216.

Flavell, J. H., Green, F. L., & Flavell, E. R. (1986). Development of knowledge about the appearance-reality distinction. *Monographs of the Society for Research in Child Development, 51*(Serial No. 212).

Fournier, M., Combès, B., Roberts, N., Braga, J., & Prima, S. (2011). Mapping the distance between the brain and inner surface of the skull and their global asymmetries. In N. J. Pelc, B. M. Dawant, R. M. Summers, K. H. Wong, D. J. Manning, W. W. Boonn, J. D'hooge, et al., (Eds.), *Medical Imaging 2011: Image Processing,* (pp. 79620Y–79620Y–7). Orlando, FA. doi:10.1117/12.876795

Fox, N. A., Rubin, K. H., Calkins, S. D., & Marshall, T. R. (1995). Frontal activation asymmetry and social competence at four years of age. *Child Development, 66,* 1770–1784.

Friedman, O., & Leslie, A. M. (2004). A developmental shift in processes underlying successful belief-desire reasoning. *Cognitive Science, 28,* 963-977.

German, T. C., & Cohen, A. S. (2011). A cue-based approach to " theory of mind ": Re-examining the notion of automaticity. *British Journal of Developmental Psychology, 30*, 45-58. doi:10.1111/j.2044-835X.2011.02055.x

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*, 26–37.

Gopnik, A., & Slaughter, V. (1991). Young children's understanding of changes in their mental states. *Child Development, 62,* 98–110.

Gopnik, A., & Wellman, H. M. (1992). Why the child's theory of mind really is a theory. *Mind and Language, 7,* 145-71.

Gopnik, A., & Wellman, H. M. (2012). Reconstructing constructivism: Causal models, Bayesian learning mechanisms, and the theory theory. *Psychological Bulletin*. doi: 10.1037/a0028044

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

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

Haeussinger, F. B., Heinzel, S., Hahn, T., Schecklmann, M., Ehlis, A.-C., & Fallgatter, A. J. (2011). Simulation of near-infrared light absorption considering individual head and prefrontal cortex anatomy: implications for optical neuroimaging. *PloS one*, *6*, e26377. doi:10.1371/journal.pone.0026377

Harris, P.L. (2006) Social cognition. In D. Kuhn & R. Siegler (Eds.) *Handbook of child psychology* (5th edition, pp. 811-858). NY: Wiley.

Huppert, T. J., Diamond, S. G., Franceschini, M. A., & David, A. (2009). HomER: A review of time-series analysis methods for near-infrared spectroscopy of the brain. *Applied Optics, 48*, D280-298.

Joanette, Y., Alice, M., Pimenta, D. M., Kristensen, C. H., & Scherer, L. C. (2008). Neuroimaging investigation of executive functions: evidence from fNIRS, *PSICO, 39,* 267–274.

Johnson, M. H., & de Haan, M. (2001). Developing cortical specialization for visual-cognitive function: The case of face recognition. In J. L. McClelland & R. S. Siegler (Eds.), *Mechanisms of cognitive development: Behavioral and neural perspectives* (pp. 253–270). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Kail, R. (1991). Developmental change in speed of processing during childhood and adolescence. *Psychological Bulletin, 109*, 490–501.

Kaiser, M. D., Hudac, C. M., Shultz, S., Lee, S. M., Cheung, C., Berken, A. M., Deen, B., et al. (2010). Neural signatures of autism. *Proceedings of the National Academy of Sciences of the United States of America*, *107*, 21223–21228. doi:10.1073/pnas.1010412107

Kalsbeek, A., Buijs, R.M., Hofman, M.A., Matthijssen, M.A., Pool, C.W., & Uylings, H.B. (1987). Effects of neonatal thermal lesioning of the mesocortical dopaminergic projection on the development of the rat prefrontal cortex. *Brain Research, 429*, 123–132.

Karmiloff-Smith, A. (1997). Crucial differences between developmental cognitive neuroscience and adult neuropsychology. *Developmental Neuropsychology*, *13*, 513–524.

Kaufman, A., Kaufman, N., 1990. Kaufman Brief Intelligence Test.

Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition*, *89*, 25-41.

Kleiner, M., Brainard, D., Pelli, D. (2007) What's new in Psychtoolbox-3? *Perception, 36 ECVP Abstract Supplement*

Kobayashi, C., Glover, G. H., & Temple, E. (2007). Children's and adults' neural bases of verbal and nonverbal "theory of mind". *Neuropsychologia*, *45*, 1522-32.

Kobayashi, C., Glover, G. H., & Temple, E. (2008). Switching language switches mind: Linguistic effects on developmental neural bases of "Theory of Mind". *Social Cognitive and Affective Neuroscience*, *3*, 62-70.

Kovelman, I., Mascho, K., Millott, L., Mastic, A., Moiseff, B., & H Shalinsky, M. (2012). At the rhythm of language: brain bases of language-related frequency perception in children. *NeuroImage*, *60*, 673–82. doi:10.1016/j.neuroimage.2011.12.066

Kovelman, I., Shalinsky, M. H., White, K. S., Schmitt, S. N., Berens, M. S., Paymer, N., & Petitto, L.-A. (2009). Dual language use in sign-speech bimodal bilinguals: fNIRS brain-imaging evidence. *Brain and language*, *109*, 112–23. doi:10.1016/j.bandl.2008.09.008

Lackner, C. L., Bowman, L. C., & Sabbagh, M. A. (2010). Dopaminergic functioning and preschoolers' theory of mind. *Neuropsychologia, 48*, 1767–1774.

Lackner, C. L., Sabbagh M. A., Hallinan, E., Liu, X., Holden J. J. (2012). Dopamine receptor D4 gene variation predicts preschoolers' developing theory of mind. *Developmental Science, 15,* 272-80. doi: 10.1111/j.1467-7687.2011.01124.x.

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

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

Lloyd-Fox, S., Blasi, A., & Elwell, C. E. (2010). Illuminating the developing brain: the past, present and future of functional near infrared spectroscopy. *Neuroscience and Biobehavioral reviews*, *34*, 269–84. doi:10.1016/j.neubiorev.2009.07.008

Malle, B.F. (2004). *How the mind explains behavior: Folk explanations, meaning, and social interaction.* Cambridge, MA: MIT Press.

Marshall, D. H., Drummey, A. B., Fox, N. A., & Newcombe, N. S. (2002). An event-related potential study of item recognition memory in children and adults. *Journal of Cognition and Development, 3,* 201–224.

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,* 1199–1208.

Miller, S. A. (2012) *Theory of mind: beyond the preschool years.* New York: Psychology Press.

Minagawa-Kawai, Y., Van der Lely, H., Ramus, F., Sato, Y., Mazuka, R., & Dupoux, E. (2011). Optical brain imaging reveals general auditory and language-specific processing in early infant development. *Cerebral cortex, 21*, 254–61. doi:10.1093/cercor/bhq082

Mitchell, J. P. (2008). Activity in right temporo-parietal junction is not selective for theory-of-mind. *Cerebral Cortex*, *18*, 262-71.

Moor, B. G., Op de Macks, Z. A., Güroglu, B., Rombouts, S. A. R. B., Van der Molen, M. W., & Crone, E. A. (2012). Neurodevelopmental changes of reading the mind in the eyes. *Social cognitive and affective neuroscience*, *7*, 44-52. doi:10.1093/scan/nsr020

Moll, H., & Tomasello, M. (2004). 12- and 18-month-old infants follow gaze to spaces behind barriers. *Developmental Science*, 7, F1-F9.

Mosconi, M. W., Mack, P. B., McCarthy, G., & Pelphrey, K. A. (2005). Taking an "intentional stance" on eye-gaze shifts: A functional neuroimaging study of social perception in children. *NeuroImage*, *27*, 247-52.

Nunez, P. L. (1995). Neocortical dynamics and human EEG rhythms. New York: Oxford University Press.

Ohnishi, T., Moriguchi, Y., Matsuda, H., Mori, T., Hirakata, M., Imabayashi, E., Hirao, K., et al. (2004). The neural network for the mirror system and mentalizing in normally developed children: An fMRI study. *NeuroReport*, *15*, 1483-1487.

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

Ozonoff, S., Pennington, B. F., & Rogers S. J (2006). Executive functioning deficits in high-functioning autistic individuals: Relationship to theory of mind. *Journal of Child Psychology and Psychiatry, 32*, 1081-1105.

Pascual-Marqui, R. D. (2002). Standardized low-resolution brain electromagnetic tomography (sLORETA): Technical details. *Methods and Findings in Experimental and Clinical Pharmacology, 24*, 5–12.

Pascual-Marqui, R. D., Esslen, M., Kochi, K., & Lehmann, D. (2002). Functional imaging with low resolution brain electromagnetic tomography (LORETA): A review. *Methods and Findings in Experimental and Clinical Pharmacology, 24,* 91–95.

Pelli, D. G. (1997) The VideoToolbox software for visual psychophysics: Transforming numbers into movies, *Spatial Vision 10,* 437-442.

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

Peterson, C. C. (2004). Theory-of-mind development in oral deaf children with cochlear implants or conventional hearing aids. *Journal of Child Psychology and Psychiatry, 45,* 1096-1106.

Peterson, C. C. , Wellman , H. M. , & Liu , D. (2005). Steps in theory of mind development for children with autism and deafness. *Child Development*, *76*, 502-517.

Pfeifer, J. H., Masten, C. L., Borofsky, L. a, Dapretto, M., Fuligni, A. J., & Lieberman, M. D. (2009). Neural correlates of direct and reflected self-appraisals in adolescents and adults: When social perspective-taking informs self-perception. *Child development*, *80*, 1016-38.

Phillips, A. T., Wellman, H. M., & Spelke, E. S. (2002). Infants' ability to connect gaze and emotional expression to intentional action. *Cognition, 85*, 53-53.

Pineda, J. A., & Hecht, E. (2009). Mirroring and mu rhythm involvement in social cognition: are there dissociable subcomponents of theory of mind? *Biological Psychology, 80,* 306–314.

Pizzagalli, D. A., Peccoralo, L. A., Davidson, R. J., & Cohen, J. D. (2006). Resting anterior cingulate activity and abnormal responses to errors in subjects with elevated depressive symptoms: A 128-channel EEG study. *Human Brain Mapping, 27*, 185–201.

Pizzagalli, D. A., Sherwood, R. J., Henriques, J. B., & Davidson, R. J. (2005). Frontal brain asymmetry and reward responsiveness. *Psychological Science, 16*, 805–813.

Poldrack, R. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends in Cognitive Sciences, 10*, 59–63.

Poldrack, R. (2007). Region of interest analysis for fMRI. *SCAN, 2,* 67-70. doi: 10.1093/scan/nsm006

Polich, J. (1993). Cognitive brain potentials. *Current Directions in Psychological Science, 2,* 175–262.

Popolo, M., McCarthy, D.M., & Bhide, P.G. (2004). Influence of dopamine on precursor cell proliferation and differentiation in the embryonic mouse telencephalon. *Developmental Neuroscience, 26*, 229–244.

Repacholi, B. M., & Gopnik, A. (1997). Early reasoning about desires: Evidence from 14- and 18-month-olds. *Developmental Psychology, 33*, 12-12.

Rosenthal, R. (1991) Meta-analytic procedures for social research (revised). Newbury Park, CA: Sage

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*, 1705-13.

Sabbagh, M.A. (2004). Understanding orbitofrontal contributions to theory-of-mind reasoning: Implications for autism. *Brain and Cognition, 55*, 209–219.

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*, 1147-62.

Sabbagh, M. A., Moulson, M. C., & Harkness, K. L. (2004). Neural correlates of mental state decoding in human adults: An event-related potential study. *Journal of cognitive neuroscience*, *16*, 415-26.

Saxe, R. 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*, 1197-1209.

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

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*, 692-699.

Saxe, R., & Wexler, A. (2005). Making sense of another mind: the role of the right temporo-parietal junction. *Neuropsychologia*, *43*, 1391-9.

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 Special Issue: Theory of Mind, 1, 284-284. doi: 10.1080/17470910601000446

Scholl, B. J., & Leslie, A. M. (1999). Modularity, development and 'theory of mind'. *Mind and Language, 14,* 131-153.

Sodian, B. (2011). Theory of mind in infancy, *Child Development Perspectives, 5*, 39–43. doi: 10.1111/j.1750-8606.2010.00152.x

Sommer, M., Döhnel, K., Sodian, B., Meinhardt, J., Thoermer, C., & Hajak, G. (2007). Neural correlates of true and false belief reasoning. *NeuroImage*, *35*, 1378-84.

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*, 123-31. Elsevier B.V.

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

Strangman, G., Culver, J. P., Thompson, J. H., & Boas, D. a. (2002). A quantitative comparison of simultaneous BOLD fMRI and NIRS recordings during functional brain activation. *NeuroImage*, *17*, 719–731. doi:10.1006/nimg.2002.1227

Sugiura, L., Ojima, S., Matsuba-Kurita, H., Dan, I., Tsuzuki, D., Katura, T., & Hagiwara, H. (2011). Sound to language: different cortical processing for first and second languages in elementary school children as revealed by a large-scale study using fNIRS. *Cerebral cortex*, *21*, 2374–93. doi:10.1093/cercor/bhr023

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

Tager-Flusberg, H., & Sullivan, K. (2000). A componential view of theory of mind: evidence from William's syndrome. *Cognition, 76,* 59–90.

Taylor, M. J., & Baldeweg, T. (2002). Application of EEG, ERP and intracranial recordings to the investigation of cognitive functions in children. *Developmental Science, 5,* 318–334.

Thatcher, R. W. (1992). Cyclic cortical reorganization in early childhood. *Brain and Cognition, 20,* 24–50.

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). New York: Guilford.

Thatcher, R. W., Walker, R. A., & Guidice, S. (1987). Human cerebral hemispheres develop at different rates and age. *Science, 236*, 1110–1113.

Tremblay, H., & Rovira, K. (2007). Joint visual attention and social triangular engagement at 3 and 6 months. *Infant Behavior & Development, 30*, 366-366.

Tucker, D.M. (1993). Spatial sampling of head electrical fields: the geodesic sensor net. *Electroencephalography and Clinical Neurophysiology, 87*, 154–163.

Van der Cruyssen, L., Van Duynslaeger, M., Cortoos, A., & Van Overwalle, F. (2009). ERP time course and brain areas of spontaneous and intentional goal inferences. *Social Neuroscience, 4,* 165–184.

Vogeley, K., Bussfeld, P., Newen, A., Herrmann, S., Happé, F., Falkai, P., Maier, W., et al. (2001). Mind reading: Neural mechanisms of theory of mind and self-perspective. *NeuroImage*, *14*, 170-81.

Wellman, H.M. (2002). Understanding the psychological world: developing a theory of mind. In U. Goswami (Ed.), *Blackwell handbook of childhood cognitive development.* (pp. 167–187). Malden, MA: Blackwell Publishing.

Wellman, H. M. (2011). Developing a theory of mind. In U. Goswami (Ed.), *The Blackwell handbook of childhood cognitive development* (2nd Edition) (pp. 258-284). New York: Blackwell.

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

Wellman, H. M., Fang, F., Liu, D., Zhu, L., & Liu, G. (2006). Scaling of theory of mind understanding in Chinese children. *Psychological Sciences*, *17*, 1075-1081.

Wellman, H. M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Development, 75,* 523-541.

Wellman, H. M., & Woolley, J. D. (1990). From simple desires to ordinary beliefs: The early development of everyday psychology. *Cognition*, *35*, 245-275.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13*, 103–128.

Woodward, A. (1998). Infants selectively encode the goal object of an actor' s reach. *Cognition, 69*, 1-34.

Zelazo, P. D. (2006). The dimensional change card sort (DCCS): A method of assessing executive function in children. *Nature Protocols, 1*, 297–301.