Neural Correlates of Children's Theory of Mind Development

David Liu
University of Michigan

Mark A. Sabbagh
Queen's University, Kingston

William J. Gehring and Henry M. Wellman
University of Michigan

Young children show significant changes in their mental-state understanding as marked by their performance on false-belief tasks. This study provides evidence for activity in the prefrontal cortex associated with the development of this ability. Event-related brain potentials (ERPs) were recorded as adults ($N = 24$) and 4-, 5-, and 6-year-old children ($N = 44$) reasoned about reality and the beliefs of characters in animated vignettes. In adults, a late slow wave (LSW), with a left-frontal scalp distribution, was associated with reasoning about beliefs. This LSW was also observed for children who could correctly reason about the characters' beliefs but not in children who failed false-belief questions. These findings have several implications, including support for the critical role of the prefrontal cortex for theory of mind development.

Human social interaction hinges on unique and sophisticated abilities to attribute unobservable mental states (beliefs, desires, intentions, etc.) to ourselves and others (Wellman, 1990). This ''theory of mind'' underlies human cooperation, deception, communication, and cultural learning. The everyday importance of theory of mind is most powerfully underscored in the case of autism, a neurodevelopmental disorder involving specific impairments in understanding of mental states (Baron-Cohen, Leslie, & Frith, 1985). The striking specificity of the social-cognitive impairments in autism has led to the conjecture that theory of mind development may be paced, at least in part, by changes to relatively specific neural circuitry. The goal of the present research was to provide developmental data about the neural correlates of one major milestone of young children's theory of mind development: the understanding of false beliefs.

In the false-belief task children are provided with scenarios such as the following: Max puts a puppy in a red box, and while he is away and not looking, the puppy moves from the red box to a blue box. When children between the ages of 3 and 6 years are asked where Max thinks the puppy is, they robustly develop from answering according to reality (the blue box) to answering according to Max's false belief (the red box), evidencing an emerging understanding that beliefs represent reality but nevertheless contrast with it (Wellman, Cross, & Watson, 2001). Additionally, the false-belief task appears to be especially difficult for individuals with autism relative to tasks involving other mental states (see, e.g., Peterson, Wellman, & Liu, 2005).

Although the neural correlates of changes in false-belief performance have not been studied in children, the question has been addressed in adult populations. Functional neuroimaging, neurophysiology, and brain lesion studies have identified a network of brain regions associated with theory of mind and thus critical to smooth navigation of the social world (Frith & Frith, 1985; Gallagher & Frith, 2003; Saxe, Carey, & Kanwisher, 2004); these regions include the medial prefrontal cortex, the temporo-parietal junction, the superior temporal sulcus, and the temporal poles. Our primary research question is whether these same brain systems are important for the development of false-belief reasoning. Specifically, do young children who perform well on false-belief tasks show evidence for activating some aspect of this network of

brain regions? Conversely, do young children who fail false-belief tasks also fail to activate this network when confronted with belief judgments?

A consistent finding in developmental cognitive neuroscience is that the neural activity associated with children's cognitive functioning is more diffuse than that of adults (Casey, Giedd, & Thomas, 2000; Johnson, 1999); increased cortical specialization seems to come with development. A secondary research question of the current study is whether the neural correlates of theory of mind follow a similar developmental trend of greater localization in adults than children.

In the current study, we recorded human event-related brain potentials (ERPs) while children 4 to 6 years of age and adults made mental-state (belief) judgments and reality judgments. This contrast between judging a person's mental state versus reality is crucial to mental-state understanding and focal to false-belief performance. Two previous ERP studies have investigated false-belief reasoning in adults (Liu, Sabbagh, Gehring, & Wellman, 2004; Sabbagh & Taylor, 2000); both studies found that false-belief reasoning was associated with a late slow wave (LSW) ERP component over left-frontal regions. Given this, the current study was designed to address two questions: (a) whether children, like adults, recruit these prefrontal regions to reason about mental states and (b) whether functional changes in the use of these prefrontal regions are associated with theory of mind development in children. Preschool-age children's prefrontal cortex certainly undergoes major changes, as evidenced, for example, in significant changes in frontal electroencephalogram (EEG) coherence (Thatcher, 1992). The current study tested the hypothesis that prefrontal regions are recruited during false-belief reasoning only by children who are generally successful at false-belief tasks. Such results would suggest strongly that the specific neural computations underlying adult mentalizing develop in conjunction with the emergence of mentalizing in young children.

Method

Participants

Twenty-nine adults participated in the study. Five participants did not provide at least 15 usable, artifact-free electrophysiological data trials for each condition and were excluded from the final sample of 24 adult participants (8 men and 16 women) for analysis. The sources of the artifact data included eye blinks, eye movements, and head and body movements. All adult participants were right-handed and had normal or corrected-to-normal vision.

Seventy 4-, 5-, and 6-year-olds participated in the study. Of the 70 children, 4 children stopped before completing at least half of the trials, and an additional 22 children's electrophysiological data did not provide at least 15 usable, artifact-free trials for each condition. Data of these 26 children were excluded (this proportion of attrition is common in ERP studies with infants and young children; DeBoer, Scott, & Nelson, 2004). The excluded group of children and the final sample of children did not differ in age, $t(68) = 0.34$, ns; in performance on the false-belief questions, $t(68) = 0.28$, ns; or in performance on the true-belief questions, $t(68) = 0.96$, ns. The final sample for analysis thus consisted of forty-four 4-, 5-, and 6-year-olds (22 boys and 22 girls; mean age = 5 years 11 months). All of the children were reported by their parents to be right-handed. All adult and child participants were recruited from a Midwestern American college town. Primarily, participants were of European-American descent. There were no known neurological or neurodevelopmental disorders in any participants.

Stimuli

We constructed a multitrial theory-of-mind task (consisting of multiple, animated false-belief scenarios) that was suitable for collecting ERPs from participants (Figure 1). Participants were presented with 40 similar animation trials; 24 trials were false-belief trials and 16 trials were true-belief trials. True-belief trials were included so that participants would pay attention to each trial. In addition, to ensure participants paid attention to each trial, a reality control question was also asked on every trial, and participants repeated any trials in which they answered the reality question incorrectly.

The structure of all 40 trials was the same, beginning with a cartoon character standing next to two boxes holding two animals. The cartoon character puts one animal in one box and the other animal in the other box and then walks in front of the boxes so that he or she cannot see either box. One of the animals in the boxes jumps out of the box and either moves to the other box (30 trials) or goes back into the same box (10 trials). After this portion of each trial (which took 15–20 s), participants were asked by the experimenter to make a reality judgment and a think judgment. Thus, each trial provides data for both the think condition and the
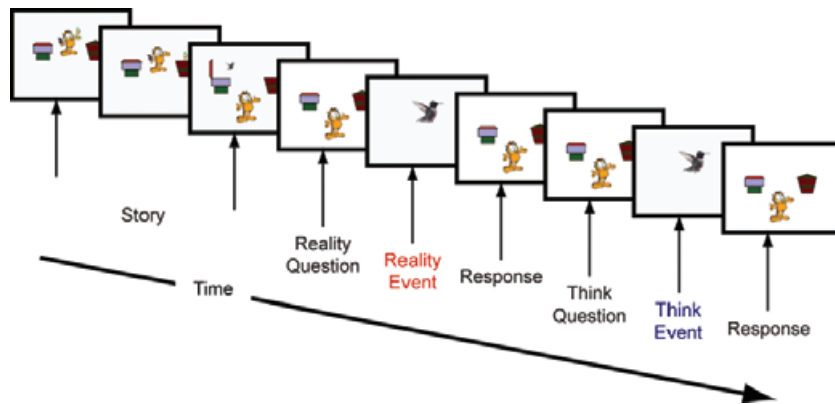
Figure 1. Multitrial theory-of-mind task.
Note. For each trial, following the story scenario, participants made a belief judgment and a reality judgment. The order of the belief and reality questions were randomized across trials. The event-related brain potentials (ERPs) for the belief condition were time-locked to the ''think'' events, and the ERPs for the reality condition were time-locked to the ''reality'' events.

reality condition. For a reality judgment, participants were asked to judge where one of the two animals was in reality (''*Really*, where is this?''), followed by the presentation of one of the two animals. This pictorial presentation of a single animal was the target event to which the reality condition ERP data were time-locked. For a think judgment, participants were asked to judge where the cartoon character thinks one of the two animals is (e.g., ''Where does Garfield *think* this is?''), followed by the presentation of one of the two animals. This pictorial presentation of a single animal was the target event to which the think condition ERP data were time-locked. When the think judgment was about an animal that moved to another box, it was a false-belief question (24 trials); when the think judgment was about an animal that stayed in the same box, it was a true-belief question (16 trials). False-belief and true-belief trials were randomly ordered, and the order of the reality and belief questions within each trial was random across trials.

Procedure

For each trial, participants were presented with the unfolding events and asked, verbally, by the experimenter to make a reality judgment and a think judgment. For the child participants, the experimenter monitored their attention to the stories and the ERP-eliciting stimuli; the experimenter made sure the child was attending to the computer screen before asking the judgment questions and triggering the ERP-eliciting stimulus. Each ERP-eliciting stimulus was presented for 2,000 ms. After the offset of the ERP-eliciting stimulus, participants provided their answers verbally or by pointing to one of the two boxes.

EEG Recording and Analysis

The 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 ERP system uses 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 to 100 Hz elliptical band-pass filter and digitized at 250 Hz sampling rate. Adult participants' continuous EEG data were segmented to epochs of 1,000 ms after stimulus onset with a 100-ms prestimulus baseline; child participants' continuous EEG data were segmented to epochs of 1,500 ms after stimulus onset with a 100-ms prestimulus baseline (a longer epoch for children because their ERP components occurred at longer latencies).

Artifacts were identified in the children's data with the following steps. For each trial, channels were marked for artifact if signal amplitude exceeded 100 μV or if a running average of activity exceeded 50 μV (this detects sharp transitions in the signal). Because children's EEG data vary immensely between individuals, subsequent to this automated process, each trial was manually inspected. Trials with more than 20 channels marked with artifact were excluded. For trials with less than 20 channels marked with artifact, an algorithm that derives values from neighboring channels via spherical spline interpolation was

used to replace bad channels. EEG data were then corrected for eye-blink and eye-movement artifacts using the Gratton, Coles, and Donchin (1983) algorithm. Artifacts were identified in the adults' data with these same steps, except with different thresholds: Channels were marked for artifact if signal amplitude exceeded 50 μV or if a running average of activity exceeded 25 μV.

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. Prior to analysis, the ERP data were corrected to the 100-ms prestimulus baseline and digitally filtered with a 30-Hz low-pass filter.

Results

Adults

Adult participants were near perfect in answering reality ($M = 100\%$), false-belief ($M = 98\%$ correct), and true-belief questions ($M = 99\%$). Based on the results of previous mentalizing ERP studies (Liu et al., 2004; Sabbagh & Taylor, 2000), we anticipated that the difference between waveforms for belief and reality judgments would show a left-frontal scalp distribution. It is clear from visual inspection of the adult grand average ERP waveforms for belief and reality judgments from a left-frontal scalp location (Figure 2, top left) that there is a late differentiation between the conditions. To confirm this, mean amplitude in the 775- to 850-ms

poststimulus epoch was computed for each condition from electrodes in a 3×3 grid encompassing scalp locations from left to right (laterality) and from anterior to posterior (caudality): F5, Fz, F6, C3, FCz, C4, PO3, POz, and PO4. A 2 (condition: belief vs. reality) $\times$ 3 (laterality) $\times$ 3 (caudality) repeated measures analysis of variance (ANOVA) was conducted on the mean amplitudes of single electrodes in the 3×3 grid of scalp locations. When necessary, for all of our analyses, p values were adjusted using the Greenhouse–Geisser correction.

Our focus was interaction effects with condition, because these tested effects associated with belief judgments versus reality judgments. There was a Condition $\times$ Caudality interaction, $F(2, 46) = 4.41$, $p = .035$, $MSE = 3.48$, $\eta_p^2 = .16$, and a Condition $\times$ Laterality interaction, $F(2, 46) = 8.53$, $p = .002$, $MSE = 3.26$, $\eta_p^2 = .27$. Most important, there was a three-way interaction between condition, caudality, and laterality, $F(4, 92) = 5.08$, $p = .006$, $MSE = 1.36$, $\eta_p^2 = .18$, revealing a larger effect of condition (belief vs. reality) from the left-frontal scalp location than from posterior and right-lateral scalp locations. Targeted comparisons between belief and reality conditions for each of the three frontal electrodes confirmed the left lateralization of the frontal effect, shown in Figure 3 (top). Thus, there was a significant difference at the left-frontal electrode, $t(23) = 3.44$, $p = .002$, but not at mid- or right-frontal sites, $t(23) = -0.58$, ns; $t(23) = -1.45$, ns respectively. In sum, confirming previous findings with adults (Liu et al., 2004; Sabbagh & Taylor,
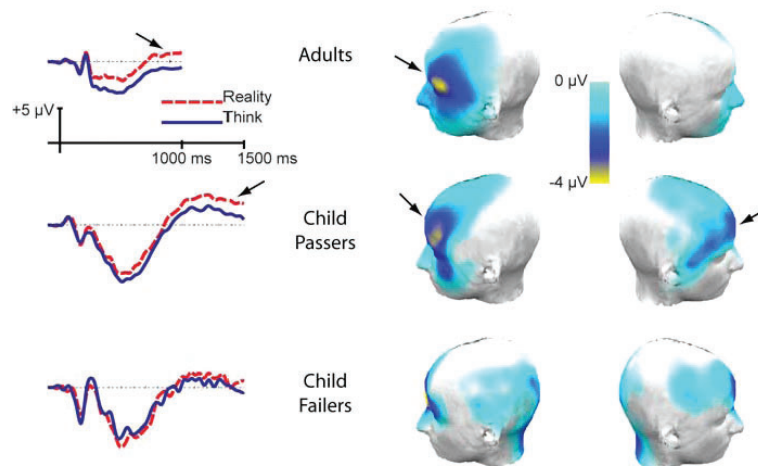

Figure 2. Event-related brain potentials (ERPs) waveforms and maps of scalp electrical activity.
Note. Left: The ERP waveforms from a left-frontal (F5) electrode for each of the three groups: adults (top), child passers (middle), and child failers (bottom). The ERP waveforms extend for 1,000 ms for adults and 1,500 ms for children. The dashed red lines indicate the reality condition, and the solid blue lines indicate the think condition. The arrows indicate the late slow wave (LSW). Right: The maps of scalp electrical activity—mean amplitude difference between conditions (reality subtracted from belief) in the 775- to 850-ms poststimulus epoch for adults (top) and in the 1,400- to 1,500-ms poststimulus epoch for child passers (middle) and child failers (bottom).
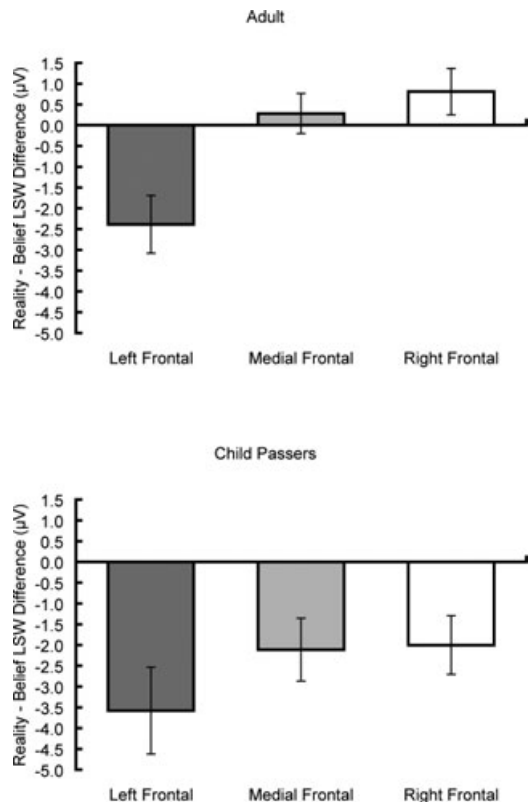
Figure 3. Frontal mean late slow wave (LSW) difference.
Note. Mean event-related brain potential (ERP) amplitude difference between conditions (reality subtracted from belief) from a left-frontal (F5), a mid-frontal (Fz), and a right-frontal (F6) electrode in the LSW for adults (top) and child passers (bottom).

2000), our results show a left-frontal scalp distribution for the LSW that differentiated belief judgments from reality judgments. This is further illustrated in the topographic map of scalp electrical activity (Figure 2, top right), which displays the mean amplitude difference between conditions (reality subtracted from belief) in the 775- to 850-ms poststimulus epoch.

Children's Behavioral Performance

Analysis of the behavioral data of the 44 children who provided usable ERP data revealed a clear bimodal distribution in performance, where most children were either consistently correct on false-belief judgments or consistently incorrect. This is in keeping with the notion that theory of mind development involves major conceptual changes (Wellman et al., 2001). To compare children based on their understanding of false belief, we classified the children as "passers" ($n = 28$) if they were consis-

tently correct (greater than 75% correct) on false-belief judgments and as "failers" ($n = 13$) if they were consistently incorrect (less than 25% correct). Three children did not show either pattern and were thus set aside from the focal analyses. Passers were slightly older on average ($M = 6$–3) than failers ($M = 5$–7), $t(39) = 2.94$, $p = .006$.

We compared the ERP data of children who were consistently correct on false-belief judgments to those of children who were consistently incorrect. We first examined children who provided at least 15 usable, artifact-free ERP trials per condition ($n = 41$) in order to analyze a larger sample of children. However, we followed these analyses by examining only the children who provided at least 25 ERP trials per condition ($n = 35$) to confirm our findings with a more reliable, conservative sample. These confirmatory analyses strengthen confidence in the findings because they go beyond simply examining data from participants who on average provided a small number of usable, artifact-free trials, as is often typical in ERP studies with infants and young children.

Child Passers

Paralleling the analyses of adults, we analyzed child ERP data with the same 3×3 grid of scalp locations. Visual inspection revealed that the morphology of child passers' ERP waveforms and adults' ERP waveforms were very similar (Figure 2, middle left). The waveforms of child passers show a similar late differentiation between the belief and reality conditions. The child passers' LSW occurred later than that of adults, consistent with a general age-related change in the latency of ERP components (DeBoer et al., 2004; Taylor & Baldeweg, 2002), which is possibly linked to the development of faster information processing throughout childhood (Kail, 1991). Mean amplitude for each condition was calculated relative to baseline in the 1,400- to 1,500-ms poststimulus epoch. A 2 (condition: belief vs. reality) $\times$ 3 (caudality) $\times$ 3 (laterality) repeated measures ANOVA was conducted on the mean amplitudes in the 3×3 grid of scalp locations.

Again, our focus was interaction effects with condition, because these contrast belief judgments versus reality judgments. There was a Condition $\times$ Caudality interaction, $F(2, 54) = 10.04$, $p = .001$, $MSE = 16.98$, $\eta_p^2 = .27$, but no Condition $\times$ Laterality interaction, $F(2, 54) = 0.34$, *ns*. Importantly there was again a three-way interaction between condition, caudality, and laterality, $F(4,$

108) = 2.79, $p = .055$, $MSE = 5.83$, $\eta_p^2 = .09$. Similar to adults, the child passers' LSW differentiated belief judgments from reality judgments and showed a left-frontal scalp distribution. This is further illustrated in the topographic map of scalp electrical activity (Figure 2, middle right). Although the difference between belief and reality conditions was greatest in the left-frontal electrode for child passers, visual inspection of their and adults' topographic maps suggests that children have a less localized, and more diffuse, scalp distribution. Targeted comparisons between conditions revealed a significant difference for all three frontal electrodes (Figure 3, bottom): left-frontal electrode, $t(27) = 3.43$, $p = .002$; mid-frontal electrode, $t(27) = 2.80$, $p = .009$; and right-frontal electrode, $t(27) = 2.84$, $p = .009$. This is in contrast to the results from adults, where only the left-frontal electrode showed a significant difference between belief and reality conditions. Thus, child passers (whose performance on false-belief judgments was similar to that of adults) also displayed a left lateralized LSW, but one that was more diffuse and less lateralized than adults.

Analysis of the more conservative sample of children who provided at least 25 ERP trials per condition yielded the identical pattern of findings. There was a two-way interaction between condition and caudality, $F(2, 48) = 8.06$, $p = .003$, $MSE = 18.48$, $\eta_p^2 = .25$, but not a significant interaction between condition and laterality, $F(2, 48) = 0.22$, ns. There was a three-way interaction between condition, caudality, and laterality, $F(4, 96) = 2.98$, $p = .053$, $MSE = 5.24$, $\eta_p^2 = .11$. Targeted comparisons between conditions revealed a significant difference for all three frontal electrodes, all $ps < .05$.

Child Failers

Next we analyzed the ERP data of the children who were consistently incorrect on false-belief judgments. The child failers' grand average ERP waveforms for belief and reality judgments from a left-frontal scalp location (Figure 2, bottom left) showed no late differentiation between the conditions, and visual inspection of the electrodes in the 3×3 grid revealed that child failers' ERP waveforms did not differentiate belief judgments from reality judgments. As with the child passers, mean amplitude for each condition was calculated relative to baseline in the 1,400- to 1,500-ms poststimulus epoch, and a 2 (condition: belief vs. reality) $\times$ 3 (caudality) $\times$ 3 (laterality) repeated measures ANOVA was conducted on the mean amplitudes. Confirming a lack of differ-

entiation between belief and reality conditions, there were no significant two-way or three-way interactions with condition, $F(4, 48) = 0.99$, ns (see also Figure 2, bottom right). Analysis of the more conservative sample of children who provided at least 25 ERP trials per condition confirmed these results: no significant two- or three-way interactions with condition, $F(4, 36) = 0.41$, ns.

Further Analyses

Child passers and failers demonstrated different patterns of results, but passers were also slightly older on average. Therefore, we compared *age-matched* subgroups of child passers and failers. Considering a narrower age window of 1 year, the ages of the two child groups overlapped most between the ages of 5 years 6 months and 6 years 6 months. Within that age range, because there were more passers than failers, each of the children in the failers group was matched with a child closest in age in the passers group, resulting in eight matched pairs. There was no significant difference in age between these two resulting subgroups of child passers ($M = 5$–11; $N = 8$) and failers ($M = 6$–0; $N = 8$), $t(14) = 0.22$, ns. A priori planned comparisons of left-frontal electrode activity were made between conditions. Passers in this subgroup showed a significant difference, $t(7) = 3.26$, $p < .05$, whereas failers did not, $t(7) = 0.61$, ns. A 2 (condition: belief vs. reality) $\times$ 2 (group: passers vs. failers) repeated measures ANOVA for the left-frontal electrode confirmed this pattern, revealing a marginally significant interaction between condition and group, $F(1, 14) = 3.66$, $p = .077$, $MSE = 6.36$, $\eta_p^2 = .21$. This direct comparison of age-matched child passers and failers, albeit with a small sample, bolsters the findings of a late differentiation in the ERP waveforms between the belief and reality conditions in child passers, but not in child failers.

Discussion

Our results demonstrate for the first time the crucial developmental emergence of neural activity associated with the ability to reason about mental states. A frontal LSW characterized adults and children who understood false belief, but not children who failed at false-belief reasoning. The frontal scalp distribution of this LSW presumably reflects the activity of the prefrontal cortex (although the spatial resolution of ERP precludes definitive identification of the cortical structures associated with the activity, even with source analysis).

Nevertheless, source analysis in a previous ERP study (Liu et al., 2004), which observed a frontal LSW with the same scalp distribution as observed in the current study, using the same exact tasks, points to activity in the prefrontal cortex for the frontal LSW. Thus, our findings suggest a critical role for the prefrontal cortex in the employment *and* in the development of theory of mind.

The contributions of prefrontal cortex to theory of mind development are further revealed in the additional developmental pattern from child passers to adults. This comparison showed increasing localization of the frontal functions associated with continuing mastery in mentalizing. Child passers displayed a pattern similar to adults' negative LSW, but these children's ERP activity had a more diffuse frontal scalp distribution, even though the child passers and adults showed similar, consistently correct, performance on false-belief questions. These results are consistent with developmental ERP research in other cognitive domains, which often observes more diffuse (and less localized) ERP activity in children than in adults performing the same task (Johnson, 1999). This is especially true of prefrontal cortical regions, which have a protracted course of postnatal development relative to other regions of the brain, as found in postmortem human neuroanatomy studies (Huttenlocher, 1990, 1997). Consequently, neuroimaging studies with children generally find diffuse, bilateral activation extending later in age in the prefrontal cortex than in other brain regions (Casey et al., 2000). It has been proposed that some neurocognitive specializations are a result of dynamic developmental interactions with the environment (Johnson, 1999, 2001; Karmiloff-Smith, 1992, 1998). Although our data do not directly address the role of interactions with the environment in such an account, they are consistent with a *developmental* story of brain function specialization and localization for mentalizing from preschool years to adulthood.

Often, researchers make statements about theory of mind as if children undergo a singular change from not having to having a theory of mind between 3 and 6 years of age, based on the rapid achievement of false-belief competence. This depiction is inaccurate in part because children progressively understand different mental states (Peterson et al., 2005; Wellman & Liu, 2004). However, our finding of diffuse ERP activity in children who comprehend false belief suggests that this depiction is inaccurate even for false-belief competence and hints at continued developmental consolidation even after children achieve consistently correct false-belief performance.

Demonstrating a left-frontal negative LSW to be a correlate of belief-reasoning in adults and children now raises a further question: What are the computational processes in the prefrontal cortex associated with the negative LSW in the context of belief-reasoning? Previous ERP research has found different forms of LSWs in connection with processes of working memory systems. For instance, LSWs are sustained in relation to the length of time working memory is activated (Ruchkin, Canoune, Johnson, & Ritter, 1995), suggesting that slow waves are a reflection of extended working-memory processing. It is argued that positive and negative slow waves reflect perceptual and conceptual memory processes, respectively (Ruchkin, Johnson, Mahaffey, & Sutton, 1988). That is, positive slow waves reflect the difficulty of perceptual operations whereas negative slow waves reflect the difficulty of conceptual operations in working memory (e.g., a left-frontal negative LSW is associated with translation of letters into Morse codes; Lang et al., 1987). Different scalp distribution of negative LSWs reflect processing of different domains of information (Mecklinger & Pfeifer, 1996; Ruchkin, Johnson, Grafman, & Canoune, 1992); right-frontal and posterior negative LSWs are associated with spatial and object working memory whereas the left-frontal negative LSW is associated with verbal working memory. Note that left-frontal negative slow waves are not associated with simple verbal recognition or recall or with episodic or source memory retrieval; these memory processes are associated with right-frontal and posterior positive slow waves (Wilding & Rugg, 1996). Instead, a left-frontal negative LSW appears to reflect more complex operations of verbal or conceptual working memory.

We propose, therefore, that the left-frontal negative LSW associated with belief-reasoning partly reflects conceptual operations in verbal working memory recruited to solve mentalizing problems. This hypothesis is consistent with the current findings, relying on standard subtraction methodology. Both belief and reality questions required participants to solve spatial problems (as noted above, spatial working memory is associated with right-frontal and posterior LSW, not left-frontal negative LSW). Reality questions simply initiate spatial working memory and do not require any mentalizing; belief questions initiate spatial working memory but more focally require social-cognitive inferential processing in verbal working memory as well.

A central question about theory of mind is whether development is a result of domain-specific or domain-general changes in the children's cognitive abilities, such as working memory (Gordon & Olson, 1998), executive functions (Moses & Sabbagh, 2007; Müller, Zelazo, & Imrisek, 2005), and language abilities (Astington & Jenkins, 1999). A possible interpretation of the left-frontal LSW is that it reflects domain-general processes, such as working memory. For this reason, we included a control comparison between judgments about belief and reality, both of which require similar spatial working memory and executive functioning; however, some domain-general differences could remain and that is a question for future research. Nevertheless, we prefer a domain-specific interpretation for the current findings. Sabbagh and Taylor (2000) compared ERP waveforms associated with false-belief reasoning versus false-photograph reasoning (both share the same domain-general demands) and found a left-frontal ERP component associated with reasoning about beliefs. We observed a similar left-frontal ERP component, using a different but equally important control contrast (reality rather than false photographs). Taken together these control contrasts add confidence to the conclusion that the LSW we demonstrate here is associated with domain-specific aspects of theory of mind.

In conclusion, theory of mind rapidly develops in the late preschool years. Critical to this development are abilities to reason about thinking in contrast to, and as distinct from, reality itself. For this reason, understanding false belief is a hallmark of a mature theory of mind and an important index of development in typically developing children as well as an index of impairments in individuals with autism. We demonstrate that the development of this crucial capacity to reason about belief versus reality is associated with neurophysiological changes in the prefrontal ERP. The developments we have charted thus inform research and theory about how social cognition and the brain develop together.

References

Astington, J. W., & Jenkins, J. M. (1999). A longitudinal study of the relation between language and theory-of-mind development. *Developmental Psychology*, *35*, 1311–1320.

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

Casey, B. J., Giedd, J. N., & Thomas, K. M. (2000). Structural and functional brain development and its relation to cognitive development. *Biological Psychology*, *54*, 241–257.

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.

Frith, C. D., & Frith, U. (1985). Interacting minds—A biological basis. *Science*, *286*, 1692–1695.

Gallagher, H. L., & Frith, C. D. (2003). Functional imaging of ''theory of mind.'' *Trends in Cognitive Sciences*, *7*, 77–83.

Gordon, A. C., & Olson, D. R. (1998). The relation between acquisition of a theory of mind and the capacity to hold in mind. *Journal of Experimental Child Psychology*, *68*, 70–83.

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

Huttenlocher, P. R. (1990). Morphometric study of human cerebral cortex development. *Neuropsychologia*, *28*, 517–527.

Huttenlocher, P. R. (1997). Regional differences in synaptogenesis in human cerebral cortex. *Journal of Comparative Neurology*, *387*, 167–178.

Johnson, M. H. (1999). Cortical plasticity in normal and abnormal cognitive development: Evidence and working hypotheses. *Development & Psychopathology*, *11*, 419–437.

Johnson, M. H. (2001). Functional brain development in humans. *Nature Reviews Neuroscience*, *2*, 475–483.

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

Karmiloff-Smith, A. (1992). *Beyond modularity*. Cambridge, MA: MIT Press.

Karmiloff-Smith, A. (1998). Development itself is the key to understanding developmental disorders. *Trends in Cognitive Sciences*, *2*, 389–398.

Lang, M., Lang, W., Uhl, F., Kornhuber, A., Deecke, L., & Kornhuber, H. H. (1987). Slow negative potential shifts in a verbal concept formation task. *Human Neurobiology*, *6*, 183–190.

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

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

Moses, L. J., & Sabbagh, M. A. (2007). Interactions between domain specific and domain general processing in the development of children's theory of mind. In M. J. Roberts (Ed.), *Integrating the mind: Domain general versus domain specific processes in higher cognition* (pp. 275–291). New York: Psychology Press.

Müller, U., Zelazo, P. D., & Imrisek, S. (2005). Executive function and children's understanding of false belief:

How specific is the relation? *Cognitive Development*, *20*, 173–189.

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.

Ruchkin, D. S., Canoune, H. L., Johnson, R., & Ritter, W. (1995). Working memory and preparation elicit different patterns of slow wave event-related brain potentials. *Psychophysiology*, *32*, 399–410.

Ruchkin, D. S., Johnson, R., Grafman, J., & Canoune, H. (1992). Distinctions and similarities among working memory processes: An event-related potential study. *Cognitive Brain Research*, *1*, 53–66.

Ruchkin, D. S., Johnson, R., Mahaffey, D., & Sutton, S. (1988). Toward a functional categorization of slow waves. *Psychophysiology*, *25*, 339–353.

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

Saxe, R., Carey, S., & Kanwisher, N. (2004). Understanding other minds: Linking developmental psychology and functional neuroimaging. *Annual Review of Psychology*, *55*, 87–124.

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 during early childhood. *Brain and Cognition*, *20*, 24–50.

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

Wellman, H. M. (1990). *The child's theory of mind*. Cambridge, MA: MIT Press.

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

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

Wilding, E. L., & Rugg, M. D. (1996). An event-related potential study of recognition memory with and without retrieval of source. *Brain*, *119*, 889–905.