



Published in final edited form as:

Child Dev. 2009 ; 80(4): 1163–1171. doi:10.1111/j.1467-8624.2009.01323.x.

Neural Correlates of Belief- and Desire-Reasoning

David Liu,

University of California, San Diego

Andrew N. Meltzoff, and

University of Washington

Henry M. Wellman

University of Michigan

Abstract

Theory of mind requires an understanding of both desires and beliefs. Moreover, children developmentally acquire an understanding of desires before beliefs, even when compared via tasks with matched formats and demands. Little is known about the mechanisms underlying this developmental lag or the neurocognitive underpinnings of desire reasoning. Previous functional neuroimaging and neurophysiological studies have neglected the direct comparison of these developmentally critical mental-state concepts. Here we provide evidence for frontal and right-posterior neural activities associated with desire- and belief-judgments. Event-related brain potentials (ERP) were recorded as adult participants performed tasks requiring reasoning about diverse-desires and diverse-beliefs (and as a control, about parallel physical situations). A late slow wave (LSW), with a mid-frontal scalp distribution, was associated with desire- and belief-judgments. A LSW, with a right-posterior scalp distribution, was only associated with belief-judgments. These findings demonstrate neural overlap as well as critical differences in reasoning explicitly about desires and beliefs and provide an explanation for the developmental lag: children need to recruit additional neural processes for reasoning about beliefs beyond a common neural system for reasoning about mental states more generally (desires and beliefs).

Central to children's social-cognition is the development of theory of mind – understanding that the actions of self and others are attributable to internal mental states, such as beliefs, desires, and intentions (Wellman, 1990). Theory of mind goes beyond social perception and involves the ability to conceptualize and reason about mental states in order to predict and explain actions. Often, theory of mind is discussed as a singular cognitive process with a singular developmental milestone, the passing of false-belief tasks (Wellman, Cross, & Watson, 2001). Investigations of false-belief understanding in behavioral studies with typically developing children, children with autism, and non-human primates, and more recently, in functional neuroimaging and neurophysiological studies have indeed proved fruitful in revealing the nature of theory of mind. However, focus on a single task-type can obscure crucial developmental aspects of theory of mind. Theory of mind involves the understanding of multiple interconnected mental concepts that perhaps develop at different ages.

Increasingly, research suggests that young children consistently acquire an understanding of some mental states before others (see meta-analysis in Wellman & Liu, 2004). Moreover, preschool-age children not only show a developmental trend when comparing pairs of

mental concepts on explicit understanding, but they follow a consistent progression for an extended sequence of developments (Harris, Rosnay, & Pons, 2005; Wellman & Liu, 2004). By far the most documented developmental progression is that children's explicit understanding of desires consistently precedes their explicit understanding of beliefs, even when compared via tasks with matched formats and demands (e.g., Bartsch & Wellman, 1995; Flavell, et al., 1990; Gopnik & Slaughter, 1991; Peterson, Wellman, & Liu, 2005; Wellman & Liu, 2004; Wellman & Woolley, 1990). For example, compare the diverse-desires versus diverse-beliefs tasks (used by Wellman & Liu, 2004; and also Wellman & Woolley, 1990). These tasks are matched on procedural methodology, linguistic structure, and materials. For the diverse-desires task, children are told about a story protagonist who had a choice of either a cookie or a carrot for snack and the children are asked which of the two snacks they themselves would *like* (e.g., the child prefers the cookie). But children hear that the protagonist in fact *likes* the other snack, the one that the child did not choose (e.g., the carrot). Then, children are asked to predict which snack the protagonist would choose. For the matching diverse-beliefs task, children are told about a story protagonist who had a choice of looking for her cat under the porch or in the garage and are asked which of the two locations they themselves *think* the cat is located (e.g., the child thinks the cat is in the garage). Children then hear that the story protagonist in fact *thinks* the cat is at the other location (e.g., under the porch). Finally, children are asked to predict which one of the locations the story protagonist would choose to look for the cat. Although, the task demands for the two tasks are matched, children consistently pass the diverse-desires task at an earlier age than the diverse-beliefs task. (Recent looking-time studies on infants' false-belief understanding (e.g., Onishi & Baillargeon, 2005) have added fuel to the debate about the age children possess the capacity to reason about beliefs, as least implicitly. Nevertheless, the data from studies of explicit performance strongly support a developmental trend of children having an explicit understanding of desires before an explicit understanding of beliefs.)

Why is there this well-established developmental lag in children's explicit understanding of desires versus beliefs? To date, the neurocognitive mechanisms underlying this developmental lag remain essentially unexplored. One potential neural explanation is that this progression results from there being different neural processors for different mental states, and these different processors have different developmental timetables. An alternative explanation could be that this progression results from a single processor for mental states that develops greater and greater capacity to process more complex mental states. Addressing this question of whether different mental states recruit different processors speaks to a broader issue of whether increasing competence in theory of mind with development is a matter of linking up disparate abilities or of later understandings building upon earlier ones. The current study addresses this issue with a neuroscientific approach. The use of functional neuroimaging or neurophysiology to define the neural circuitries associated with processing different mental states can potentially allow for inferences about multiple processors versus a single processor.

Functional neuroimaging studies have identified a network of brain regions associated with theory of mind (Frith & Frith, 1985; Gallagher & Frith, 2003; Saxe, Carey, & Kanwisher, 2004); these regions include the medial prefrontal cortex (PFC), the temporo-parietal junction (TPJ), the superior temporal sulcus (STS), and the temporal poles. Human event-related brain potential (ERP) studies have found a late slow wave (LSW) component over frontal regions associated with false-belief reasoning (Liu, Sabbagh, Gehring, & Wellman, 2004, in press; Sabbagh & Taylor, 2000). In almost all of these studies of the neural correlates of theory of mind, the participants were asked to reason about false belief, or simply mental states in general, in comparison to reasoning about physical phenomena. Saxe and Powell (2006) compared reasoning about beliefs with reasoning about a different, specific internal (but not mental) state – bodily sensations. They found that activations in the

left and right temporo-parietal junction as well as the posterior cingulate were selectively associated with reasoning about beliefs but not reasoning about bodily sensations. In an fMRI study primarily examining a careful contrast between mental reasoning about false beliefs versus physical reasoning about false photographs, Saxe and Kanwisher (2003) also included mental stories about desires. They briefly reported that TPJ regions involved in false-belief reasoning in contrast to physical reasoning also responded significantly to desire reasoning. The data for processing desires specifically were not mentioned or analyzed further, but their graphical results suggest that the neural response to false-belief stories was decidedly greater than the neural response to desire stories. Thus there are almost no data available on the neural correlates of reasoning about desires versus beliefs, and none comparing responses to strictly matched tasks, yet this seems particularly important. And to reiterate, the distinction between understanding/attributing desires versus beliefs is one of the most well-documented developmental findings in the literature on theory of mind. In the current study we recorded ERPs while participants made diverse-desires, diverse-beliefs, and (as a control) diverse-physical judgments. The tasks were modeled after the ones used in behavioral studies that have found consistent developmental lags in young children's understanding of well-matched tasks focusing on desires versus beliefs (Peterson et al., 2005; Wellman & Liu, 2004; Wellman & Woolley, 1990).

Examining the neural substrates of reasoning about desires as well as beliefs is important not only for developmental issues and accounts. By most analyses, both philosophical (e.g. Churchland, 1984; Dennett, 1987; Stich, 1983) and psychological (Wellman, 1990; Premack & Woodruff, 1978; Baron-Cohen, 1995), an everyday folk psychology depends on a system of belief-desire reasoning, in which attributions of both these mental states are crucial for human understanding of actions and minds. Here we address understanding of both desires and beliefs and the neural substrates involved in reasoning about them.

Methods

Participants

Twenty-four adults (Mean age = 22 years; 14 males and 10 females) participated in the study. Nine participants did not provide at least 20 usable, artifact-free electrophysiological data trials for each condition and were excluded from the final sample of 15 participants (10 males and 5 females) for ERP analyses; the sources of the artifact data included eye blinks, eye movements, and head and body movements. All participants were right handed and had normal or corrected-to-normal vision.

Stimuli and procedure

We constructed multi-trial diverse-desires, diverse-beliefs, and diverse-physical judgment tasks that were suitable for collecting ERPs from participants. Participants were presented with 48 trials for each condition: Desires, Beliefs, and Physical. The structure of all 48 trials of all three conditions was the same. In each trial, participants were first provided 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. The participants read the information (e.g., the boy likes grapes, but the girl likes celery) accompanied by pictures. On a random third of the trials this was followed by a memory check to ensure participants paid attention to each trial. If the participants answered the memory question incorrectly, the information phase of the trial was repeated. After the information phase of each trial (which took 7300 ms), the participants read the target question (details provided below for each condition) and were then 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 the ERP data were time-locked. The participants then answered the target question.

For the Desires condition, in each trial, participants read about a closed box said to contain either a snack or a toy and a boy who likes a particular food or toy (e.g., grapes) and a girl who likes a different food or toy (e.g., celery). Whether a boy or a girl was presented first was counter-balanced. Then, when the story was about food, the participants read one of two target questions about what would happen when the box was opened: “Who says ‘I want some’ when they see this?” or “Who says ‘I won’t have any’ when they see this?” When the story was about toys, the participants read one of two target questions: “Who says ‘I’ll play with it’ when they see this?” or “Who says ‘I won’t play with it’ when they see this?” The two questions (one positive and one negative) for each type (food or toys) were chosen randomly in each trial. After the target question, the participants were presented with a picture of one of the two food/toys (e.g., celery). Note that, in this and other conditions, the participants were not able to answer the target question until after being presented with a picture of the food/toy that was in the closed box. After 2000 ms of seeing the revealed food/toy, the participants provided their answer by choosing one of the two characters.

For the Beliefs condition, in each trial, participants read about a mystery box that contained food/toys for a guessing game. The participants read about a girl who thinks the mystery box contains a particular food or toy (e.g., blocks) and a boy who thinks the mystery box contains a different food or toy (e.g., markers). Whether a boy or a girl was presented first was counter-balanced. Then, the participants read one of two target questions: “Who says ‘I was right’ when they see this?” or “Who says ‘I was wrong’ when they see this?” The two questions were chosen randomly in each trial. After the target question, the participants were presented with a picture of one of the two food/toys (e.g., blocks). The participants then provided the answer by choosing one of the two characters.

The Physical condition provided a non-mental control condition. For the Physical condition, in each trial, participants read about a closed box that contained food/toys to put away. The participants read about a red bin that should hold a particular food or toy (e.g., blocks) and a blue bin that should hold a different food or toy (e.g., markers). Then, the participants read one of two target questions: “Where do you put this?” or “Where do you not put this?” The two questions were chosen randomly in each trial. After the target question, the participants were presented with a picture of one of the two food/toys (e.g., markers) the one that was in the closed box. The participants then provided the answer by choosing one of the two bins. Note that the trials for all three conditions were constructed to have the same perceptual structure. The participants could take a low-level strategy of simply matching each food/toy with an entity (character or bin). Therefore, any differences between conditions would point to the mental-state processing beyond these perceptual and task similarities.

The trials for each condition and type (e.g., Desires condition about toys) were presented in blocks of six trials. The blocks semi-randomly alternated between the conditions, with the stipulation that no condition and type repeated successively.

Electrophysiological recording and analysis

The electroencephalogram (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 K Ω (the 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 Hz to 100 Hz elliptical bandpass filter and digitized at 250 Hz sampling rate. Continuous EEG

data were segmented to epochs of 1000 ms after stimulus onset with a 100 ms pre-stimulus baseline.

Artifacts were identified in the EEG data with the following steps. For each trial, channels were marked for artifact if a running average of activity exceeded 40 μV (this detects sharp transitions in the signal). Subsequent to this automated process, each trial was manually inspected. Trials with more than 15 channels marked with artifact were excluded. For trials with less than 15 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. 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 pre-stimulus baseline and digitally filtered with a 30 Hz low-pass filter.

Results

As expected, the participants' performances on the diverse-beliefs, diverse-desires, and diverse-physical judgment tasks were near perfect (96.0%, 96.4%, and 96.5% correct respectively). The performances were well matched and did not differ between the conditions: Beliefs versus Physical, $t(23) = 0.74$, *ns*; Desires versus Physical, $t(23) = 0.14$, *ns*; Beliefs versus Desires, $t(23) = 0.82$, *ns*.

We designed the physical condition to provide a control baseline. Therefore, any differences in waveforms between the Beliefs and Physical conditions reveals the components associated with diverse-beliefs judgments, and any differences in waveforms between the Desires and Physical conditions reveals the components associated with diverse-desires. Any difference in waveforms between the Beliefs and Desires conditions reveals the components that differ between diverse-beliefs and diverse-desires judgments. Figure 1 displays the grand average waveforms for all three conditions from electrodes in a 3 \times 3 grid encompassing scalp locations from left to right (laterality) and from anterior to posterior (caudality): F5, Fz, F6, C5, Cz, C6, P5, Pz, and P6. When necessary, for all of our analyses, *p*-values were adjusted using the Greenhouse-Geisser correction.

It is clear from visual inspections of the waveforms in Figure 1 that there is a late mid-frontal differentiation between the Beliefs (more positive) and Physical conditions and that there is also a late mid-frontal differentiation between the Desires (more positive) and Physical conditions. These differentiations are around 800 ms post-stimulus. To confirm this, mean amplitude in the 800-850 ms post-stimulus epoch was computed for each condition from all electrodes in the 3 \times 3 grid of scalp locations. Visual inspections of the waveforms in Figure 1 also suggest that there is an effect of condition in the right-posterior location around 600-800 ms post-stimulus. To confirm this, mean amplitudes in the 600-700 ms and 700-800 ms post-stimulus epochs were computed for each condition from all electrodes in the 3 \times 3 grid of scalp locations. In addition, to check whether there were earlier components, mean amplitudes in the 500-600 ms post-stimulus epoch were also computed for each condition from all electrodes in the 3 \times 3 grid of scalp locations.

Frontal component

Based on results of previous ERP studies on reasoning about beliefs (e.g., Liu et al., 2004; Sabbagh & Taylor, 2000) we predicted that the diverse-beliefs judgment component, and perhaps the diverse-desires judgment component, would yield a late slow wave with frontal scalp distributions. An omnibus 3 (condition) \times 3 (laterality) \times 3 (caudality) repeated measures ANOVA was conducted on the mean amplitudes of single electrodes in the 3 \times 3

grid of scalp locations for the 800-850 ms post-stimulus epoch. The focus of our analyses is on interaction effects with condition, since these tested components associated with diverse-beliefs, diverse-desires, and physical judgments. For the 800-850 ms epoch there was, importantly, a significant three-way interaction between condition, caudality, and laterality, $F(8, 112) = 2.65, p = .03, MSE = 0.87$; this was the only interaction effect with condition. The three-way interaction confirms what is observed in the waveforms in Figure 1, a late mid-frontal differentiation between the Beliefs (more positive) and Physical conditions and a late mid-frontal differentiation between the Desires (more positive) and Physical conditions.

To target the interaction effect with condition, we focused on the frontal electrodes and analyzed these condition contrasts: Beliefs versus Physical, Desires versus Physical, and Beliefs versus Desires. For each contrast, a 2 (conditions: Beliefs vs. Physical, Desires vs. Physical, or Beliefs vs. Desires) \times 3 (laterality) repeated measures ANOVA was conducted on the mean amplitudes of the frontal electrodes for the 800-850 ms epoch with follow-up *t*-tests of single locations.

There was a significant interaction effect of Beliefs versus Physical and laterality, $F(2, 28) = 5.24, p = .01, MSE = 0.83$, and there was a significant interaction effect of Desires versus Physical and laterality, $F(2, 28) = 5.72, p = .02, MSE = 0.68$. However, there was not a significant interaction effect of Beliefs versus Desires and laterality or a significant main effect of Beliefs versus Desires, $F(2, 28) = 0.44, ns$, and $F(2, 14) = 1.54, ns$, respectively. Examination of individual locations revealed, at the mid-frontal location, a significant effect of Beliefs versus Physical, $t(14) = 3.38, p < .01$, and a significant effect of Desires versus Physical, $t(14) = 3.05, p < .01$. There were no differences between the Beliefs and Physical conditions or between the Desires and Physical conditions at the left-frontal location or at the right-frontal location. Finally, there was not a significant effect of Beliefs versus Desires at any of the three frontal locations, confirming the ANOVA findings. This pattern of results suggests that in the 800-850 ms epoch, there is a mid-frontal ERP component associated with *both* diverse-beliefs judgments and diverse-desires judgments. This is further illustrated in the topographic map of scalp electrical activity shown in Figure 2 (bottom for the 800-850 ms epoch), which displays the mean amplitude difference of each contrast (Physical subtracted from Beliefs, Physical subtracted from Desires, and Desires subtracted from Beliefs). Darker red indicates greater positive difference between conditions. For the 800-850 ms epoch, the difference between Beliefs and Physical and the difference between Desires and Physical both show a mid-frontal scalp distribution, which is not observed in the difference between Beliefs and Desires.

Additionally, to check whether there were frontal components in earlier epochs, a 2 (conditions: Beliefs vs. Physical, Desires vs. Physical, or Beliefs vs. Desires) \times 3 (laterality) repeated measures ANOVA was conducted on the mean amplitudes of the frontal electrodes for the 500-600 ms, 600-700 ms, and 700-800 ms post-stimulus epochs. For each of these epochs, there were no main or interaction effect of any of the condition contrasts. In sum, the results show a positive late slow wave (LSW) with a mid-frontal scalp distribution associated with judgments about diverse beliefs and diverse desires.

Posterior component

An omnibus 3 (condition) \times 3 (laterality) \times 3 (caudality) repeated measures ANOVA was conducted on the mean amplitudes of single electrodes in the 3 \times 3 grid of scalp locations for the 600-700 ms and the 700-800 ms post-stimulus epochs. For the 600-700 ms epoch, there were no significant interactions with condition; for the 700-800 ms epoch, there was a moderately significant three-way interaction between condition, caudality, and laterality, $F(8, 112) = 2.27, p = .065, MSE = 0.53$. This confirms what is observed in the waveforms in Figure 1 – a late mid-frontal differentiation between the Beliefs (more positive) and Physical

conditions and a late mid-frontal differentiation between the Desires (more positive) and Physical conditions.

To examine more closely the effects of condition, we focused on the posterior electrodes and analyzed these condition contrasts: Beliefs versus Physical, Desires versus Physical, and Beliefs versus Desires. For the 600-700 ms epoch, examination of individual locations revealed, at the right-posterior location, a significant effect of Beliefs versus Physical, $t(14) = 3.41, p < .01$. There were no differences between the Beliefs and Physical conditions at the left-posterior location or at the mid-posterior location, and there were no differences between the Desires and Physical conditions at any of the posterior locations. For the 700-800 ms epoch, examination of individual locations revealed, at the right-posterior location, a significant effect of Beliefs versus Physical, $t(14) = 4.29, p < .001$, and a significant effect of Beliefs versus Desires, $t(14) = 2.37, p = .03$. There were no differences between the Beliefs and Physical conditions or between the Beliefs and Desires conditions at the left-posterior location or at the mid-posterior location; there were no differences between the Desires and Physical conditions at any of the posterior locations. This pattern of results suggests that starting in the 600-700 ms post-stimulus epoch and more robustly in the 700-800 ms post-stimulus epoch, there is a right-posterior ERP component associated with diverse-beliefs judgments, but not diverse-desires judgments. This is further illustrated in the topographic map of scalp electrical activity show in Figure 2 (top for the 600-700 ms epoch), which displays the mean amplitude difference of each contrast (Physical subtracted from Beliefs, Physical subtracted from Desires, and Desires subtracted from Beliefs). For the 600-700 ms epoch, the difference between Beliefs and Physical and the difference between Beliefs and Desires both show a right-posterior scalp distribution, which is not observed in the difference between Desires and Physical. In sum, the results show a positive late slow wave (LSW) with a right-posterior scalp distribution associated with judgments about diverse beliefs only. Finally, there were no significant differences between the conditions at any of the three central locations for any of the epochs. Thus, the effects of conditions were only in frontal and posterior locations.

Discussion

The results of the current study show the network of neural circuitry associated with reasoning specifically about beliefs and desires. Previous functional neuroimaging and neurophysiological studies of theory of mind have not focused on the *direct* comparison of these developmentally critical mental-state concepts. We found that the neural circuitries associated with processing different mental states overlap in some ways, but not, critically, in other ways. A mid-frontal late slow wave (LSW) was observed for explicit belief judgments (consistent with previous ERP studies focused on false beliefs; Liu, Sabbagh, Gehring, & Wellman, 2004, in press; Sabbagh & Taylor, 2000) and also in our data for explicit desire judgments. However, a right-posterior LSW (perhaps occurring slightly earlier) was observed for explicit belief judgments only.

A mature, adult-like theory of mind requires an understanding of both desires and beliefs. By most analyses, both philosophical (e.g. Churchland, 1984; Dennett, 1987; Stich, 1983) and psychological (Wellman, 1990; Premack & Woodruff, 1978; Baron-Cohen, 1995), an everyday naïve psychology is centered by a system of belief-desire reasoning, in which attributions of both these mental states are crucial for human understanding of actions and minds. In addition, a well-established developmental progression, at least for explicit reasoning tasks of the sort used with children of preschool age and older, is that children consistently demonstrate an understanding of desires at an earlier age than an understanding of beliefs, even when tested via tasks with matched formats and demands (e.g., Bartsch & Wellman, 1995; Flavell, et al., 1990; Gopnik & Slaughter, 1991; Peterson, Wellman, & Liu,

2005; Wellman & Liu, 2004; Wellman & Woolley, 1990). The current ERP study provides evidence for different networks of neural circuitries associated with this reasoning about beliefs versus desires. Whereas the processing of diverse-beliefs involved both the frontal and the right-posterior neural systems, the processing of diverse-desires was observed to recruit only the frontal system. This suggests that reasoning about these different mental states involve different sorts of processes.

The pattern of results for the frontal LSW and the posterior LSW potentially addresses findings in the functional neuroimaging literature on the medial prefrontal cortex (PFC), the temporo-parietal junction (TPJ), and the superior temporal sulcus (STS; Frith & Frith, 1985; Gallagher & Frith, 2003; Saxe, Carey, & Kanwisher, 2004). The predominant approach has been to determine which of these neural regions are specifically associated with theory of mind. However, our results suggest that the question should not be about theory of mind as a singular process, but should be about which aspect of theory of mind is subserved by which network of brain systems. In the current study, we addressed just two aspects of theory of mind: diverse-desires and diverse-beliefs judgments. Our data suggest that the right-posterior system is involved in processing diverse-beliefs but not diverse-desires, and this is consistent with arguments and data indicating the TPJ as the critical region for false-belief understanding in adults (e.g., Saxe, Carey, & Kanwisher, 2004). Our ERP data are less spatially precise than is possible with fMRI, but are nonetheless consistent with such findings. Complementarily, the frontal system, apparent in our data and in other data as well, now appears to be more broadly associated with more general aspects of thinking about the mind, including desires and beliefs. It should be noted that although we observed the frontal system processing beliefs and desires and the right-posterior system processing beliefs, the processing of these two mental states could recruit yet more other neural systems that were not identified with an ERP approach. Similarly, our tasks asked participants to reason about the diversity of desires, which did not recruit the posterior system, but other aspects of reasoning about desires could still involve the posterior system (e.g., in Saxe & Kanwisher, 2003). Additionally, our tasks asked participants to reason about the diversity of beliefs, and thus differ from the previous ERP studies with false-belief tasks (Liu, Sabbagh, Gehring, & Wellman, 2004, in press; Sabbagh & Taylor, 2000). Although we found a similar frontal LSW, the ERP studies with false-belief tasks showed a slightly more left-frontal scalp distribution than our findings. Therefore, the current study suggests a more fine-tuned analysis of ERP and functional neuroimaging mentalizing studies is needed for various specific mental states and the types of reasoning associated with the specific mental states. The neural basis of theory of mind can no longer be thought of as a singular process.

The pattern of results for diverse-desires and diverse-beliefs judgments suggests a possible explanation for the developmental lag in children demonstrating an explicit understanding of desires at an earlier age than an understanding of beliefs. Intriguingly, it is possible that children need to develop just one system for diverse-desires but need to develop both systems for diverse-beliefs. It appears that belief judgments overlap the neural system capable of processing desire judgments, but require the involvement of an additional system. This suggests a developmental account where children's understanding of beliefs builds upon their understanding of desires. That is, on this account children first have some understanding of the diversity of desires, and the understanding of the diversity of beliefs builds on that earlier understanding by involving the same mental-state processing characteristic of desires plus an additional belief-specific processing as well. Hence the behavioral data consistently show children understanding desires before they are capable of understanding beliefs. Still further research is required with children to test whether this is indeed an explanation for the development of explicit understanding of desires and beliefs. These data and arguments apply most directly to the explicit reasoning of the sort targeted in standard belief-desire reasoning tasks. It is an open question altogether as to the neural

substrates that support implicit social-cognitive reasoning of the sort targeted in recent theory of mind research with infants (Onishi & Baillargeon, 2005).

Our data provide a direct comparison between the neural substrates for reasoning about desires as well as beliefs. The procedures not only compare these two mental states, they yield telling findings. Beliefs and desires, as mental states, involve similar processing in frontal regions of cortex. However, beliefs additionally recruit right posterior cortical processing as well, processing not evident for desires. The findings are important in their own right for the light they shed on neural correlates of mental-state processing and for their implications for the neural mechanisms that shape the development of these processes. Moreover, the methods developed here are suitable for children (and now validated with adults). We are currently engaged in research with these methods with children. This will allow us to tackle directly the developmental course for reasoning about these two critical mental states.

Acknowledgments

This research was supported by funds awarded to A. Meltzoff from the National Science Foundation (SBE-0354453) as part of the LIFE Science of Learning Center and to H. Wellman from the National Institutes of Health (HD-22149). Any opinions, findings, and conclusions expressed in the paper are those of the authors and do not necessarily reflect the views of NSF.

References

- Baron-Cohen, S. *Mindblindness: An essay on autism and theory of mind*. MIT Press; Cambridge, MA: 1995.
- Bartsch, K.; Wellman, HM. *Children talk about the mind*. Oxford University Press; New York: 1995.
- Churchland, PM. *Matter and consciousness: A contemporary introduction to the philosophy of mind*. The MIT press; Cambridge, MA: 1984.
- Dennett, DC. *The intentional stance*. The MIT press; Cambridge, MA: 1987.
- Flavell JH, Flavell ER, Green FL, Moses LJ. Young children's understanding of fact beliefs versus value beliefs. *Child Development* 1990;61:915–928. [PubMed: 2209196]
- Frith CD, Frith U. Interacting minds—a biological basis. *Science* 1985;286:1692–1695. [PubMed: 10576727]
- Gallagher HL, Frith CD. Functional imaging of 'theory of mind'. *Trends in Cognitive Sciences* 2003;7:77–83. [PubMed: 12584026]
- Gopnik A, Slaughter V. Young children's understanding of changes in their mental states. *Child Development* 1991;62:98–110.
- Gratton G, Coles MG, Donchin E. A new method for off-line removal of ocular artifact. *Electroencephalography & Clinical Neurophysiology* 1983;55:468–484. [PubMed: 6187540]
- Harris PL, Rosnay M, Pons F. Language and children's understanding of mental states. *Current Directions in Psychological Science* 2005;14:69–73.
- Liu D, Sabbagh MA, Gehring WJ, Wellman HM. Decoupling beliefs from reality in the brain: an ERP study of theory of mind. *NeuroReport* 2004;15:991–995. [PubMed: 15076721]
- Liu D, Sabbagh MA, Gehring WJ, Wellman HM. Neural Correlates of Children's Theory of Mind Development. *Child Development*. in press.
- Onishi KH, Baillargeon R. Do 15-month-old infants understand false beliefs? *Science* 2005;308:255–258. [PubMed: 15821091]
- Peterson CC, Wellman HM, Liu D. Steps in theory of mind development for children with autism and deafness. *Child Development* 2005;76:502–517. [PubMed: 15784096]
- Premack D, Woodruff G. Does the chimpanzee have a theory of mind? *Behavioral and brain sciences* 1978;1:515–526.
- Sabbagh MA, Taylor M. Neural correlates of the theory-of-mind reasoning: An event-related potential study. *Psychological Science* 2000;11:46–50. [PubMed: 11228842]

- Saxe R, Powell L. Its the thought that counts: Specific brain regions for one component of theory of mind. *Psychological Science* 2006;17:692–699. [PubMed: 16913952]
- Saxe R, Carey S, Kanwisher N. Understanding other minds: Linking developmental psychology and functional neuroimaging. *Annual Review of Psychology* 2004;55:87–124.
- Stitch, S. *From folk psychology to cognitive science*. Bradford Books; Cambridge, MA: 1983.
- Tucker DM. Spatial sampling of head electrical fields: The geodesic sensor net. *Electroencephalography and Clinical Neurophysiology* 1993;87:154–163. [PubMed: 7691542]
- Wellman, HM. *The child's theory of mind*. MIT Press; Cambridge, MA: 1990.
- Wellman HM, Liu D. Scaling of theory-of-mind tasks. *Child Development* 2004;75:523–541. [PubMed: 15056204]
- Wellman HM, Woolley JD. From simple desires to ordinary beliefs: The early development of everyday psychology. *Cognition* 1990;35:245–275. [PubMed: 2364653]
- Wellman HM, Cross D, Watson J. Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development* 2001;72:655–684. [PubMed: 11405571]

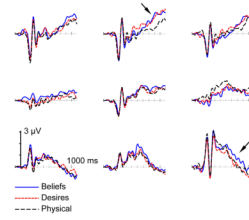


Figure 1. Grand average ERP waveforms for the Desires (red dotted lines), Beliefs (blue solid lines), and Physical (black dotted lines) conditions from nine electrodes in a 3×3 grid encompassing scalp locations from left to right and front to back (top to bottom in the grid). The arrows indicate the frontal LSW and the right-posterior LSW.

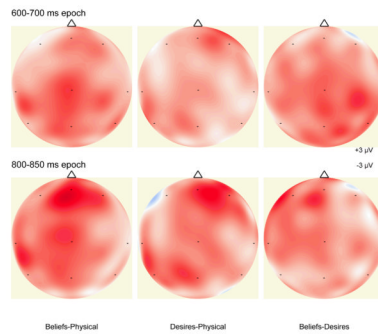


Figure 2. Maps of the scalp electrical activity – mean amplitude difference of each contrast between conditions (Physical subtracted from Beliefs, Physical subtracted from Desires, and Desires subtracted from Beliefs) in the 600-700 ms post-stimulus epoch (top) and in the 800-850 ms post-stimulus epoch (bottom). The maps are oriented with frontal up from a view from above the scalp.