Automatic Theory of Mind: Covert Detection via Brain Activity

David Liu and Joe Dietzel

University of California, San Diego

Author Note

Correspondence concerning this article should be addressed to David Liu, Department of Psychology, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0109. E-mail: davidliu@ucsd.edu.

Word Count: 4,676

Abstract

The present research examined whether people automatically compute others' mental states and predict their actions. We leveraged the event-related brain potential (ERP) P300 component as a covert measure of automatic processing. Participants observed characters' actions in false-belief scenarios (Experiment 1) and in true-belief scenarios (Experiment 2). Participants' task was to identify when characters were at a particular spatial location; for the task, it was unnecessary to consider their mental states, and there were never any overt or implied instructions to do so. On a small proportion of the trials, the characters' movement was inconsistent with their beliefs—and only if one automatically computes the characters' mental states and predict their actions, would one process belief-violating trials as distinct from belief-consistent trials. Indeed, in Experiments 1 and 2, the belief-violating trials elicited a frontal P300 for both false- and true-belief scenarios, which suggests that participants engaged in automatic theory of mind even though they were actively engaged in another task.

Keywords: theory of mind; mentalizing; automatic processing; event-related potential (ERP)

Automatic Theory of Mind: Covert Detection via Brain Activity

1. Introduction

Humans possess the ability to reason about mental states in order to predict others' actions, and this ability—theory of mind—is fundamental to social interaction, understanding, and communication (Wellman, 1990). Nevertheless, despite accounts of mentalizing playing a central role in everyday social interaction and communication, little is known about the conditions in which people actually reason about others' mental states and actions (Apperly et al., 2006; Cohen & German, 2009). Do people reason about mental states and actions only when overtly instructed to do so or when they intentionally pause to ponder what another person thinks and wants? Or do people automatically think about others' mental states whenever they observe others' actions? One possibility is that reasoning about mental states is solely a deliberate, intentional process. Another possibility, however, is that people automatically parse others' mental states, and this is done reflexively, even when people do not have a goal to predict or explain others' actions. A third possibility is that there are two separate systems for theory of mind (Apperly & Butterfill, 2009; Sabbagh, Moulson, & Harkness, 2004; Sodian, Thoermer, & Metz, 2007)—one for deliberate and flexible reasoning about mental states and another for effortless processing of others' mental states and actions. The aim of the present research is to investigate whether people have an automatic system for computing others' mental states and predicting actions.

Apperly et al. (2006) first addressed the question of automatic mentalizing by examining response times to instructed versus incidental belief questions. In their study, participants were presented multiple trials of video stimuli depicting people in false-belief scenarios. For example, in a typical scenario, a female actor observes a male actor hide an object in one of two containers, and when she leaves the room, the male actor moves the object to the other container. One group of participants were instructed to keep track of where the female actor thinks the object is hiding, but another group of participants were instructed to keep track of where the object is hiding in reality. On a small proportion of trials, participants in both groups were randomly probed (surprised) with either a belief question about where the female actor thinks the object is hiding or a reality question about where the object is hiding. The results showed that participants instructed to keep track of beliefs answered reality probe questions as fast as belief probe questions. However, participants instructed to keep track of reality answered belief probe questions slower than reality probe questions. Apperly et al. (2006) argued this asymmetrical pattern of results suggests that the participants instructed to keep track of beliefs automatically encoded reality information, but the participants instructed to keep track of reality did not automatically encode belief information. In contrast, Cohen and German (2009) proposed an alternative interpretation of Apperly et al.'s (2006) results by noting that the elapsed time between belief cues and belief questions was longer than the elapsed time between reality cues and reality questions. By modifying the task to shorten the elapsed time between belief cues and belief questions, Cohen and German (2009) showed that response times for belief questions were equal regardless of whether participants were instructed to track beliefs or not. Thus, Cohen and German (2009) showed that the participants did automatically infer beliefs.

There are, however, some potential alternative explanations of Cohen and German's (2009) results. In their study, participants were tasked with responding to multiple trials (12 test and 48 filler trials). Thus, although participants were not overtly instructed to track beliefs, they might have realized after the first belief question that they should attend to beliefs for subsequent trials. That is, the study had an *implied instruction* to track beliefs. The possibility of implied instructions is in line with Back and Apperly's (2010) distinction between automatic and spontaneous belief inference (which can be prompted by contextual factors to comprehend the stimuli). Moreover, although participants were not provided with overt instructions, they were probed with *overt questions* and made *overt responses* about beliefs. The concern is that the belief inferences were ad hoc, after having been prompted by belief questions, and indeed, Back and Apperly (2010) presented evidence for such a concern. Therefore, it is unknown whether the Cohen and German's (2009) participants automatically inferred beliefs as they observed people's actions. What is needed to test this question is a *covert measure* of online automatic processing that does not include any instructions or questions about mental states.

Some recent studies have used more covert measures of online automatic mentalizing. Although deployed with slightly different procedures, Kovacs et al. (2010) and Samson et al. (2010) both examined whether there is automatic intrusion of someone else's visual perspective into one's own visual perspective. Both studies showed that participants' responses to what they observed were automatically influenced by what another person observed. These measures of automatic intrusion tapped into Level 1 perspective taking, which considers whether someone has perceptual access to a stimulus (Flavell, Everett, Croft, & Flavell, 1981; Samson et al., 2010). However, it remains unclear whether people automatically compute the specific contents of others' mental states to predict their actions. Examining automatic action predictions is important to address the concern of ad hoc inference of mental states.

An eye-tracking study by Senju et al. (2009) examined automatic action predictions; participants observed an actor with a false belief about to reach for a hidden object in one of two containers. The results showed that typically-developing adults looked predictively to where the actor would reach for the object, whereas adults with Asperger syndrome did not. However, because the critical test trial was preceded by several familiarization trials (in which participants were trained to expect the actor to reach through one of two windows whenever the windows were illuminated along with a chime sound), the procedure had an *implied instruction* to predict the actor's reach. Thus, following Back and Apperly's (2010) distinction, it appears that the typically-developing adults in Senju et al. (2009) engaged in spontaneous belief inference, but it is unclear whether they engaged in automatic belief inference. Nevertheless, the observed differences between typically-developing adults and adults with Asperger syndrome in Senju et al. (2009) are still informative, because whereas the typically-developing adults were cued to the procedure's implied instructions to predict the actor's reach, the adults with Asperger syndrome were not. To address the guiding question of the present research, we designed the current event-related brain potential (ERP) experiments such that the procedure was without any implied instructions to process others' mental states to predict their actions. Because our participants were instructed to actively engage in a physical identification task, the present experiments addressed the concern that previous studies (e.g., Cohen & German, 2009; Senju et al., 2009), included contextual factors that prompted implied instructions to processes others' mental states.

1.1. Using ERP as a Covert Measure

As Cohen and German (2009) noted, measures of brain activity are potentially useful tools for examining online automatic processing. For the current experiments, we developed a novel approach to covertly detect automatic mentalizing by adapting the "brain fingerprinting" protocol for ERP (Farwell & Donchin, 1991). This ERP lie-detection protocol was designed to determine whether someone is concealing guilty information. The protocol's reliability in detecting concealed guilty information at the individual level is hotly debated (see Rosenfeld, 2005); however, for our purposes, the protocol can reliably detect covert processing at the group level (Campanella, et al., 2002; Meijer, et al., 2007; Rosenfeld, Biroschak, & Furedy, 2006).

The lie-detection protocol leverages the P300 component of the ERP, which is elicited by rare, unexpected events (Farwell & Donchin, 1991). The P300 is most commonly studied with the oddball paradigm: participants are presented with a *standard* stimulus (e.g., a red circle) occurring in a large proportion of trials (e.g., 80%) and a target stimulus (e.g., a blue circle) occurring in a small proportion of trials (e.g., 20%), which results in the targets eliciting a positive potential around 300-400 ms post-stimulus (Polich & Kok, 1995). Modifying the oddball paradigm, the lie-detection paradigm adds to 20% targets and 60% standards a third stimulus category of 20% crime-relevant *probes*. It is assumed that if a person is guilty, the 20% crime-relevant probes would be meaningful to the person and would stand out as a separate category of rare, unexpected stimuli, thereby eliciting a P300. However, if a person is innocent, the crime-relevant probes would not stand out and would be viewed as being in the same category as the 60% standards, thereby not eliciting a P300. That is, for both the guilty and innocent, targets will elicit a P300 and standards will not; however, only for the guilty will probes elicit a P300. In a sense, the paradigm covertly detects whether people process probe events as distinctly separate from standard events.

2. Experiment 1

We adapted the lie-detection paradigm into a mentalizing-detection paradigm. Participants viewed multiple animation clips of people's actions. In each animation clip, the person puts an animal into either the box on the right or the box on the left and leaves the room. While the person is out of the room and cannot see, the animal moves from one box to the other (thus, the person has a false belief about the location of the animal). The person comes back to the room and walks up to one of the boxes. In target trials (20%), the person ends up at the box on the *right*, which happens to be where she believes the animal is hiding. In standard trials (60%), the person ends up at the *left* box, which also happens to be where she believes the animal is hiding. Lastly, in belief-violation (BV) probe trials (20%), the person ends up at the *left* box, but this is *inconsistent* with where she believes the animal is hiding. Participants were instructed to identify trials when the character ends up at the right box (i.e., target trials). Paralleling the logic of the lie-detection paradigm, if people are "innocent" of mentalizing, the BV probes would not stand out from the standards (in both, the character ends up at the left box), thereby not eliciting a P300. However, if people are "guilty" of mentalizing, the BV probes would stand out from the standards as rare, unexpected events, thereby eliciting a P300. That is, without any overt or implied instructions to process others' mental states to predict their actions, belief-violating actions (probe trials) would appear rare and unexpected, relative to standard trials, only if participants automatically inferred the characters' beliefs and predicted their actions.

Participants were *not* given any instructions to consider what the characters believe or what they can or cannot see; participants were *not* even given any instructions about the characters' goals, to find the animal or otherwise. As presented, participants were tasked with identifying events that occur in a particular physical location (when a character is physically behind the right box). To accomplish the task, participants could focus on the end of each trial and ignore the character and animal's actions prior to that. Thus, any mentalizing would be the result of automatic engagement from observing people's actions.

2.1. Method

2.1.1. Participants

Twenty-one adults (mean age = 20 years; range = 18 – 22 years; 11 males and 10 females) participated in the study. An additional nine participants were excluded because of experimenter errors and excessive noise in the ERP data. All participants were right handed and had normal or corrected-to-normal vision. Participants were recruited from a sample of undergraduate students and received course credit; participants were approximately 31% White, not Latino, 17% White, Latino, 3% African-American, and 49% Asian.

2.1.2. Stimuli and Procedure

We created multiple animation clips of people's actions. In each animation clip, a character puts an animal into either the right or left box and leaves the room. While the character is out of the room and cannot see, the animal moves from one box to the other. The character comes back to the room and walks up to one of the boxes. For variety, the animation clips consisted of all possible combinations of six characters (3 boys and 3 girls), two animals (a frog and a bunny), and four rooms. At the end of each animation clip, as the character starts to walk toward the boxes, the screen fades to black for 800 ms, and then, in the final image, the character is shown already standing at one of the boxes. When the screen fades to black, it is unknown which side the character will move towards—only the final image shows whether the character ends up at the box on the right or left. We did not present continuous movement to the box so that we could time-lock the ERP data to the final discrete image that revealed the ending. All animation clips lasted 5800 ms each.

Participants were randomly presented with 360 trials: 72 *target* trials (20%), 216 *standard* trials (60%), and 72 *belief-violation probe* trials (20%). For the target condition, the animal moves from the right to the left box and the character ends up at the *right* box, consistent with where he or she falsely believes the animal is hiding. For the standard condition, the animal moves from the left to the right box and the character ends up at the *left* box, consistent with where he or she falsely believes the animal is hiding. For the BV probe condition, the animal moves from the right to the left box and the character ends up at the *left* box, but this is *inconsistent* with where he or she believes the animal is hiding.

Participants were instructed to observe the actions and to press a response button when the character ends up at the right box (i.e., target trials). They were not provided with *any* instructions about the nature of the actions or the characters' mental states.

2.1.3. Electrophysiological Recording and Analysis

The electroencephalogram (EEG) was recorded continuously from scalp electrodes using HydroCel Geodesic Sensor Nets (EGI Inc., Eugene, OR), each a network of 64 Ag/AgCl electrodes embedded in an elastic geodesic tension structure. Impedance for all electrodes was kept below 40 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 – 100 Hz elliptical bandpass filter and digitized at 250 Hz sampling rate. Continuous EEG data were digitally filtered with a 40 Hz low-pass filter and segmented into epochs of 600 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 (which detects sharp transitions in the signal). 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.

Following previous P300 studies (Polich & Kok, 1995), midline electrode sites were targeted for analysis: Fz, FCz, Cz, Pz, and POz. These five electrodes encompass frontal to posterior scalp locations. We measured P300 by computing mean amplitude in the 300-400 ms post-stimulus epoch for each condition from the five electrodes. When necessary, for all of our analyses, *p*-values were adjusted using the Greenhouse-Geisser correction.

2.2. Results

Participants' performance in correctly identifying target trials was high (*M* = 98%; range = 85 – 100%). Figure 1a displays the grand average ERP waveforms of the three conditions at each of the five midline electrodes. An omnibus 3 (Condition: target, standard, and BV probe) x 5 (Scalp Location) repeated measures ANOVA was conducted on mean P300 amplitude. The results showed a significant main effect of condition, $F(2, 40) = 22.33$, $p < .001$, $\eta_p^2 = .53$, a significant main effect of scalp location, $F(4, 80) = 10.11$, $p = .002$, $\eta_p^2 = .34$, and a significant interaction between condition and scalp location, $F(8, 160) = 11.00$, $p = .001$, $\eta_p^2 = .36$. However, the focus of our analyses was not a comparison of all three conditions. Instead, the mentalizing-detection paradigm was designed to compare target versus standard conditions and, most importantly, to compare BV probe versus standard conditions. We conducted a 2 (Condition: target versus standard) x 5 (Scalp Location) repeated measures ANOVA on mean P300 amplitude. The results showed a significant main effect of condition, $F(1, 20) = 31.23$, $p < .001$, $\eta_p^2 = .61$, a significant main effect of scalp location, $F(4, 80) = 14.37$, $p < .001$, $\eta_p^2 = .42$, and a significant interaction between condition and scalp location, $F(4, 80) = 10.33$, $p = .001$, η_p^2

= .34. These results confirmed what is shown in Figure 1a: the target condition elicited a greater P300 than the standard condition, especially towards posterior scalp locations.

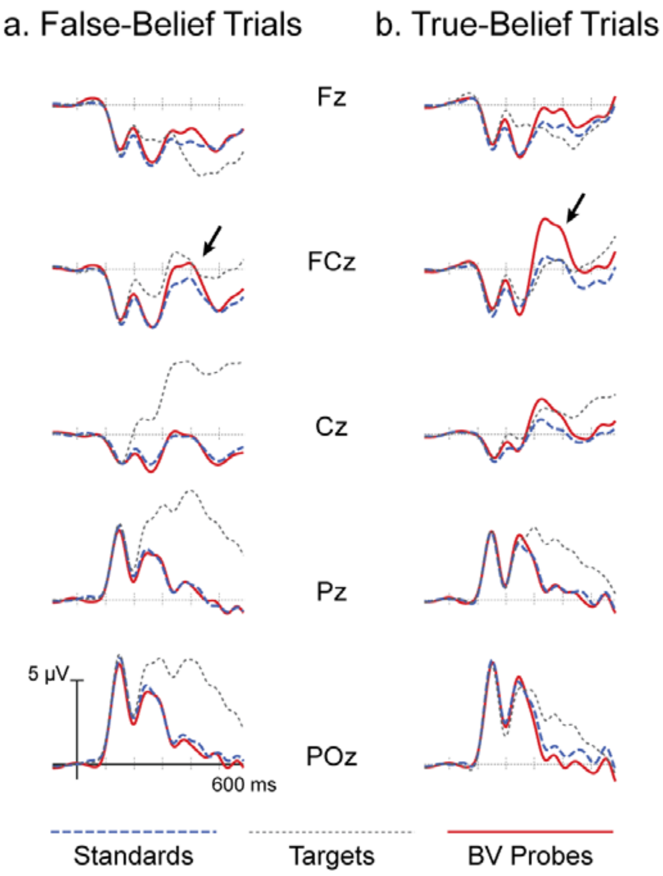

Figure 1. Grand average ERP waveforms for standard, target, and belief-violation probe conditions from five midline electrodes (Fz, FCz, Cz, Pz, and POz) for (a) false-belief trials in Experiment 1 and (b) true-belief trials in Experiment 2. From top to bottom, the five electrodes encompass frontal to posterior scalp locations. The arrows indicate the frontal P300 of greater amplitude for belief-violation probe than standard trials.

For the primary analysis, we conducted a 2 (Condition: BV probe versus standard) x 5 (Scalp Location) repeated measures ANOVA on mean P300 amplitude; Figure 2a displays the mean P300 amplitude of BV probe and standard conditions at each of the five electrodes. Although the main effect of condition was not significant, *F*(1, 20) = 2.75, *p* = .113, η_p^2 = .12, there was, *crucially*, a significant interaction between condition and scalp location, *F*(4, 80) = 3.76, *p* = .041, η_p^2 = .16. There was a marginal main effect of scalp location, *F*(4, 80) = 3.42, *p* = .056, η_p^2 = .15. Examination of individual frontal electrodes revealed a significant effect of BV probe versus standard conditions at Fz, *t*(20) = 3.56, *p* = .002, and at FCz, *t*(20) = 2.73, *p* = .013. These results confirmed what is shown in Figures 1a and 2a: the BV probe condition elicited a greater frontal P300 than the standard condition.

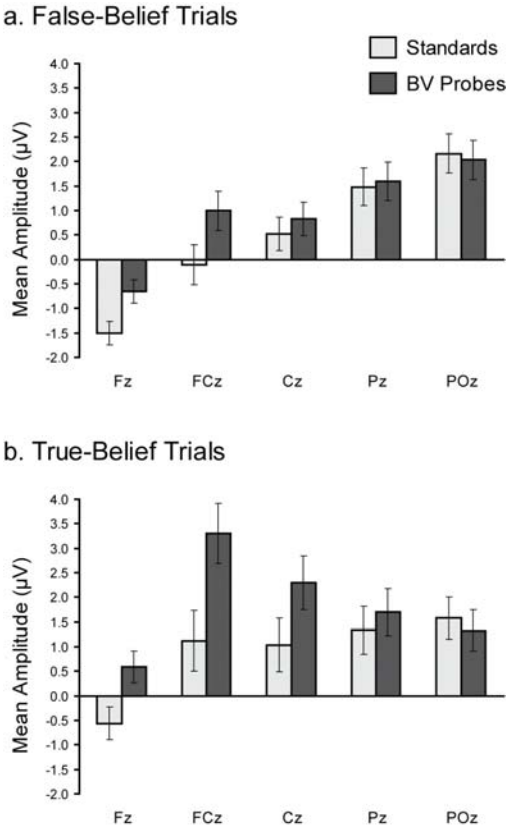

Figure 2. Mean P300 amplitude for standard and belief-violation probe conditions from five midline electrodes (Fz, FCz, Cz, Pz, and POz) for (a) false-belief trials in Experiment 1 and (b) true-belief trials in Experiment 2. From left to right, the five electrodes encompass frontal to posterior scalp locations.

2.3. Discussion

The results from Experiment 1 showed that participants automatically engaged in online mentalizing—computing the characters' false beliefs and predicting their actions. The belief-violating probe trials elicited a significantly greater frontal P300 than the belief-consistent standard trials. Following the logic of the lie-detection paradigm, the results suggest that participants processed belief-violating actions as distinctly separate from belief-consistent actions without having been prompted by any overt or implied instructions to consider the character's mental states to predict their actions.

The current results extend previous evidence of an automatic system for computing mental states and predicting actions (Cohen & German, 2009; Kovacs et al., 2010; Samson et al., 2010; Senju et al., 2009). Experiment 1 addressed the issues raised above about previous studies (Apperly et al., 2006; Cohen & German, 2009; Senju et al., 2009) by using a covert measure and eliminating implied instructions to process mental states. Participants were never given any instructions to consider the characters' goals, beliefs, or perceptual access, and they were never even asked to judge the real or believed locations of the hidden animal.

3. Experiment 2

Although the results from Experiment 1 provide substantial evidence of automatic belief inference, it is possible that such processes are restricted to false-belief scenarios, which might be especially appropriate (i.e., rare and unexpected) for eliciting a P300. In everyday life, people's beliefs are usually true, and false-belief scenarios are uncommon. Additionally, the expectation is that people with false beliefs will be surprised when they find out about reality.

Therefore, for Experiment 2, we created true-belief versions of the animation clips to examine whether automatic belief inference is specific to false-belief scenarios or more general, to include (the more common) true-belief scenarios.

3.1. Method

3.1.1. Participants

Twenty-three adults (mean age = 21 years; range = 18 – 24 years; 10 males and 13 females) participated in the study. An additional five participants were excluded because of experimenter errors and excessive noise in the ERP data. All participants were right handed and had normal or corrected-to-normal vision. Participants were recruited from a sample of undergraduate students and received course credit; participants were approximately 31% White, not Latino, 17% White, Latino, 3% African-American, and 49% Asian.

3.1.2. Stimuli and Procedure

We created true-belief animation clips so that they are exactly matched with the false-belief animation clips described in Experiment 1 *except* for the key difference of whether the character sees the animal change locations. In the true-belief version, the person comes back to the room *before* the animal moves from one box to the other (thus, the person has a true belief about the location of the animal). Other than that key difference in the animation sequence, the mentalizing-detection paradigm in Experiment 2 was the same as in Experiment 1. Additionally, electrophysiological recording and analysis were the same as in Experiment 1.

3.2. Results

Participants' performance in correctly identifying target trials was high (*M* = 91%; range = 74 – 100%). Figure 1b displays the grand average ERP waveforms of the three conditions at each of the five midline electrodes for true-belief trials. An omnibus 3 (Condition: target, standard, and BV probe) x 5 (Scalp Location) repeated measures ANOVA was conducted on mean P300 amplitude. The results showed a significant main effect of condition, $F(2, 44) = 8.58$, $p = .001$, $\eta_p^2 = .28$, a significant main effect of scalp location, $F(4, 88) = 6.43$, $p = .004$, $\eta_p^2 = .23$, and a significant interaction between condition and scalp location, $F(8, 176) = 15.52$, $p < .001$, $\eta_p^2 = .41$. To compare target versus standard conditions, we conducted a 2 (Condition: target versus standard) x 5 (Scalp Location) repeated measures ANOVA on mean P300 amplitude. The results showed a significant main effect of condition, $F(1, 22) = 15.68$, $p = .001$, $\eta_p^2 = .42$, a significant main effect of scalp location, $F(4, 88) = 11.74$, $p < .001$, $\eta_p^2 = .35$, and a significant interaction between condition and scalp location, $F(4, 88) = 12.86$, $p < .001$, $\eta_p^2 = .37$. These results confirmed what is shown in Figure 1b: the target condition elicited a greater P300 than the standard condition, especially towards posterior scalp locations.

For the primary analysis, we conducted a 2 (Condition: BV probe versus standard) x 5 (Scalp Location) repeated measures ANOVA on mean P300 amplitude; Figure 2b displays the mean P300 amplitude of BV probe and standard conditions at each of the five electrodes. Importantly, the results showed a significant main effect of condition, $F(1, 22) = 7.35$, $p = .013$, $\eta_p^2 = .25$, and a significant interaction between condition and scalp location, $F(4, 88) = 5.89$, $p = .005$, $\eta_p^2 = .21$. There was a marginal main effect of scalp location, $F(4, 88) = 3.11$, $p = .060$, $\eta_p^2 = .12$. Examination of individual frontal electrodes revealed a significant effect of BV probe versus standard conditions at Fz, $t(22) = 3.49$, $p = .002$, and at FCz, $t(22) = 3.57$, $p = .002$. These results confirmed what is shown in Figures 1b and 2b: the BV probe condition elicited a greater frontal P300 than the standard condition.

3.2.1. Comparing False-Belief and True-Belief Trials

As an exploratory analysis, we compared the frontal P300 observed with false-belief trials (Experiment 1) to the frontal P300 observed with true-belief trials (Experiment 2). We conducted a 2 (Version: false- and true-belief) x 2 (Condition: BV probe versus standard) x 2 (Frontal Scalp Locations: Fz and FCz) repeated measures ANOVA on mean P300 amplitude. The results showed a significant main effect of condition, $F(1, 42) = 23.94$, $p < .001$, $\eta_p^2 = .36$, a significant main effect of scalp location, $F(1, 42) = 25.65$, $p < .001$, $\eta_p^2 = .38$, and a significant interaction between condition and scalp location, $F(1, 42) = 5.53$, $p = .023$, $\eta_p^2 = .12$. There was not a significant main effect of version (false- versus true-belief) and none of the two- or three-way interactions with version were significant (all *p*s > .15). In summary, the results from Experiments 1 and 2 showed a frontal P300 that was greater for the BV probe condition than the standard condition for both false- and true-belief scenarios.

3.3. Discussion

The results from Experiment 2 replicate and extend the results from Experiment 1. Because the paradigm developed for the current research examines the P300 component of the ERP (which is elicited by rare, unexpected events), we conducted Experiment 2 to check that the findings from Experiment 1 are not restricted to false-belief scenarios. False-belief scenarios are themselves rare and unexpected in everyday life, and they might be especially appropriate for eliciting a P300. Nevertheless, the results from Experiment 2 showed that participants automatically engaged in online mentalizing for true-belief scenarios. The belief-violating probe trials elicited a significantly greater frontal P300 than the belief-consistent standard trials. Thus, it appears that the findings from the current research are not specific to false-belief scenarios, but are more general, to include true-belief scenarios.

4. General Discussion

Across Experiments 1 and 2, we leveraged ERP as a covert measure in a mentalizing-detection paradigm that did not include any overt or implied instructions to consider the character's mental states to predict their actions. For both true- and false-belief scenarios, the belief-violating probe trials elicited a significantly greater P300 than the belief-consistent standard trials at frontal scalp locations. That is, the participants automatically processed belief-violating actions as distinctly separate from belief-consistent actions. The current research's findings provide substantial evidence of automatic online mentalizing and prediction of actions.

Our results are consistent with previous evidence of an automatic system for computing mental states and predicting actions (Cohen & German, 2009; Kovacs et al., 2010; Samson et al., 2010; Senju et al., 2009). Nevertheless, the present findings extend previous findings by showing automatic belief inference without any implied instructions to process mental states (Apperly et al., 2006; Cohen & German, 2009; Senju et al., 2009). Furthermore, the present findings suggest that people do not engage in automatic mentalizing only in an ad hoc manner (Cohen & German, 2009; Kovacs et al., 2010; Samson et al., 2010), but rather engage in automatic online mentalizing to predict others' actions.

The present and previous findings converge on the conclusion that people have an automatic system for theory of mind. Such a conclusion opens several avenues of inquiry. In developmental research, a major unanswered question is how infants' implicit understanding of mental states (Baillargeon, Scott, & He, 2010; Senju et al., 2011; Sodian & Thoermer, 2008; Southgate, Senju, & Csibra, 2007) is associated with older children's explicit understanding of mental states (Wellman, Cross, & Watson, 2001; Wellman & Liu, 2004). One possibility is that the implicit and explicit understandings are both a part of the same theory-of-mind system

(Baillargeon, Scott, & He, 2010). However, another possibility is that they are two separate theory-of-mind systems (Apperly & Butterfill, 2009; Sabbagh, Moulson, & Harkness, 2004; Sodian, Thoermer, & Metz, 2007), with infants' implicit understanding reflecting the automatic action-prediction system and older children's explicit understanding reflecting a deliberate and flexible system for reasoning about mental states.

Further research is also needed to examine the precise contexts in which people automatically engage in theory of mind. Is it when people observe any biological action, any human action, or only certain types of human action? Another avenue for investigation is whether different systems of theory of mind are associated with different neural systems that have separate neurodevelopmental trajectories. Some possibilities are that the different systems of theory of mind are localized to different regions of the brain (Saxe, Carey, & Kanwisher) or that they have different neural timing (Liu, Meltzoff, & Wellman, 2009; Liu, Sabbagh, Gehring, & Wellman, 2004, 2009).

Lastly, the present research demonstrates the feasibility of using ERP as a covert measure of automatic processing. Farwell and Donchin's (1991) lie-detection paradigm can be adapted to covertly measure whether other cognitive processes are automatic, without any implied instructions or overt responses.

Acknowledgements

This research was supported by a grant from the National Institutes of Health to David Liu (HD066029). We thank the research assistants in the Brain and Cognitive Development Lab for assistance with data collection.

References

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

Apperly, I. A., Riggs, K. J., Simpson, A., Chiavarino, C., & Samson, D. (2006). Is Belief Reasoning Automatic? *Psychological Science, 17*(10), 841-844.

Back, E., & Apperly, I. A. (2010). Two sources of evidence on the non-automaticity of true and false belief ascription. *Cognition, 115*(1), 54-70.

Baillargeon, R., Scott, R. M., & He, Z. (2010). False-belief understanding in infants. *Trends in Cognitive Sciences, 14*(3), 110-118.

Campanella, S., Gaspard, C., Debatisse, D., Bruyer, R., Crommelinck, M., & Guerit, J. M. (2002). Discrimination of emotional facial expressions in a visual oddball task: An ERP study. *Biological Psychology, 59*(3), 171-186.

Cohen, A. S., & German, T. C. (2009). Encoding of others' beliefs without overt instruction. *Cognition, 111*(3), 356-363.

Farwell, L. A., & Donchin, E. (1991). The truth will out: Interrogative polygraphy ("lie detection") with event-related brain potentials. *Psychophysiology, 28*(5), 531-547.

Flavell, J. H., Everett, B. A., Croft, K., & Flavell, E. R. (1981). Young children's knowledge about visual perception: Further evidence for the Level 1–Level 2 distinction. *Developmental Psychology, 17*(1), 99-103.

Gratton, G., Coles, M. G., & Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography & Clinical Neurophysiology, 55*(4), 468-484.

Kovacs, A. M., Teglas, E., & Endress, A. D. (2010). The social sense: susceptibility to others' beliefs in human infants and adults. *Science, 330*(6012), 1830-1834.

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

Liu, D., Sabbagh, M. A., Gehring, W. J., & Wellman, H. M. (2004). Decoupling beliefs from reality in the brain: An ERP study of theory of mind. *NeuroReport: For Rapid Communication of Neuroscience Research, 15*(6), 991-995.

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

Meijer, E. H., Smulders, F. T. Y., Merckelbach, H. L. G. J., & Wolf, A. G. (2007). The P300 is sensitive to concealed face recognition. *International Journal of Psychophysiology, 66*(3), 231-237.

Polich, J., & Kok, A. (1995). Cognitive and biological determinants of P300: An integrative review. *Biological Psychology, 41*(2), 103-146.

Rosenfeld, J. P. (2005). 'Brain Fingerprinting': A Critical Analysis. *Scientific Review of Mental Health Practice, 4*(1), 20-37.

Rosenfeld, J. P., Biroschak, J. R., & Furedy, J. J. (2006). P300-based detection of concealed autobiographical versus incidentally acquired information in target and non-target paradigms. *International Journal of Psychophysiology, 60*(3), 251-259.

Sabbagh, M. A., Moulson, M. C., & Harkness, K. L. (2004). Neural Correlates of Mental State Decoding in Human Adults: An Event-related Potential Study. *Journal of Cognitive Neuroscience, 16*(3), 415-426.

Samson, D., Apperly, I. A., Braithwaite, J. J., Andrews, B. J., & Bodley Scott, S. E. (2010). Seeing it their way: Evidence for rapid and involuntary computation of what other people see. *Journal of Experimental Psychology: Human Perception and Performance, 36*(5), 1255-1266.

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

Senju, A., Southgate, V., Snape, C., Leonard, M., & Csibra, G. (2011). Do 18-month-olds really attribute mental states to others?: A critical test. *Psychological Science, 22*(7), 878-880.

Senju, A., Southgate, V., White, S., & Frith, U. (2009). Mindblind eyes: An absence of spontaneous theory of mind in Asperger syndrome. *Science, 325*(5942), 883-885.

Sodian, B., & Thoermer, C. (2008). Precursors to a theory of mind in infancy: Perspectives for research on autism. *The Quarterly Journal of Experimental Psychology, 61*(1), 27-39.

Sodian, B., Thoermer, C., & Metz, U. (2007). Now I see it but you don't: 14-month-olds can represent another person's visual perspective. *Developmental Science, 10*(2), 199-204.

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

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

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

Wellman, H. M., & Liu, D. (2004). Scaling of Theory-of-Mind Tasks. *Child Development, 75*(2), 523-541.

Figure Legends

Figure 1. Grand average ERP waveforms for standard, target, and belief-violation probe conditions from five midline electrodes (Fz, FCz, Cz, Pz, and POz) for (a) false-belief trials in Experiment 1 and (b) true-belief trials in Experiment 2. From top to bottom, the five electrodes encompass frontal to posterior scalp locations. The arrows indicate the frontal P300 of greater amplitude for belief-violation probe than standard trials.

Figure 2. Mean P300 amplitude for standard and belief-violation probe conditions from five midline electrodes (Fz, FCz, Cz, Pz, and POz) for (a) false-belief trials in Experiment 1 and (b) true-belief trials in Experiment 2. From left to right, the five electrodes encompass frontal to posterior scalp locations.