Young Children’s Understanding of False Belief: P3b Responses to an Implicit False Belief Task

by

Samantha Drover

A thesis submitted to the Graduate Program in Psychology in conformity with the requirements for the Degree of Master of Science

Queen’s University

Kingston, Ontario, Canada

August, 2014

Copyright © Samantha Drover, 2014
Abstract

Representational theory of mind (TOM) is the everyday understanding that others’ behaviour is motivated by internal representational mental states such as beliefs, desires and intentions. In the standard false belief task, which traditionally indexes representational TOM, a character places an object in one location, and then leaves the room. During her absence, a second character moves the object to a new location. The first character returns, and children are asked where she will look for the object. Children pass this task (answering that she’ll look in the first location) around 4–5 years. Recently, studies have found that infants and young children succeed on implicit versions of the false belief task – their anticipatory gaze and looking time align with an understanding of beliefs. Some suggest that infants and young children possess a representational TOM, masked by performance limitations, whereas others hold that the early competence relies on a non-representational, perhaps statistical basis. To address this debate, I investigated young children’s false belief understanding using a component of the event-related potential: P3b, a positive potential elicited in response to unexpected events and thought to reflect conceptual updating. Children aged 3.5 to 4.5 years and adults watched 48 trials following the standard false belief task, and P3b was measured in response to seeing the character search in belief-congruent or incongruent locations. I found a significant age group by outcome interaction: adults’ P3b responses corresponded with an understanding of others’ beliefs, whereas young children’s P3b responses reflected a lack of belief understanding. These results support the conclusion that young children do not have a representational understanding of false belief, and that perhaps their early performance on implicit false belief tasks is mediated by an early-developing non-representational understanding. These results have implications for the development of theory of mind.
Acknowledgements

First, I would like to thank my supervisor, Dr. Mark Sabbagh for his mentorship. Mark’s guidance gave me the opportunity to develop many valuable skills: clear writing, critical thinking especially with respect to statistics, and effective scientific communication, to name a few. I greatly appreciate his academic support throughout my time at Queen’s. I would also like to thank the members of my thesis committee, Dr. Valerie Kuhlmeier and Dr. Randall Flanagan, for their input and guidance throughout my Master’s. In addition to the members of my committee, I greatly appreciate the input and suggestions of all the faculty and graduate students who attended the weekly developmental proseminar meetings.

Second, thank you to my fellow graduate students in the Early Experience Lab for creating such a wonderful environment within which to learn and develop: Martha Bailey, Nicole Bardikoff, Jeanette Benson, Anastasia Christopher, Haykaz Mangardich, and Vanessa Schell.

Third, it would not have been possible to complete my Master’s project with the team of excellent undergraduate research assistants who volunteered their time to help me collect and manage the data: Rebecca Firman, Sandra Krause, Meagan Murphy, Morgan Sherrer, and Sasha Usyatynsky.

Fourth, thank you to all the families and undergraduate students who participated in my study – without their interest and support, this research would not have been possible.

Finally, I would like to extend my gratitude to my family and loved ones for your persistent interest, support, and encouragement.
# Table of Contents

Introductory Material

List of Tables vi
List of Figures vii
List of Abbreviations vii
Glossary ix

Main Body

Chapter I: Introduction 1
Chapter II: Methods 12
  A. Participants 12
  B. Materials 12
    a. Stimuli 12
    b. False Belief Battery 14
    c. Executive Functioning Battery 15
  C. EEG/ERP Recording and Data Processing 17
  D. Procedure 18
    a. Arrival 18
    b. ERP Collection 18
    c. False Belief and Executive Functioning Batteries 19
  E. Data Cleaning and Processing 19
Chapter III: Results 20
  A. Participant Attrition 20
a. Child Participants 20
b. Adult Participants 22

B. Behavioural Results 22
a. Inter-Rater Reliability 22
b. Subscale Raw Scores 22
c. Correlations 23

C. ERP Findings 24
a. Sample-by-sample paired-samples t-tests 25
c. Comparing Adults’ and Children’s P3b Responses 30
d. Predictors of the Early P3b Effect in Young Children 34

Chapter IV: Discussion 38
A. Performance on Implicit False Belief Tasks 38
B. Predictors of the P3b Amplitude in Young Children 41
C. Implications for the Development of Theory of Mind 43
D. P3b Topography and the Neural Correlates of Theory of Mind 46
E. Limitations 46
F. Conclusions 47

Additional Material
References 49
Appendix A: FB Battery Scripts and Scoring 58
Appendix B: EF Battery Scripts and Scoring 62
Appendix C: Ethics Approval and Amendments 71
# List of Tables

| Table 1. | Attrition and dropped cases for child participants | 21 |
| Table 2. | Comparisons of child participants who did and did not complete the EEG session: EF, FB battery scores, and age | 21 |
| Table 3. | Raw EF and FB battery scores | 23 |
| Table 4. | Pearson correlations among age, FB, and EF battery scores | 24 |
| Table 5. | Pearson correlations among P3b amplitudes in contiguous electrodes for adult participants | 31 |
| Table 6. | Pearson correlations among P3b amplitudes in contiguous electrodes for child participants | 32 |
| Table 7. | Pearson Correlations among P3b amplitude differences in contiguous electrodes for child participants | 35 |
| Table 8. | Regression results: Predicting amplitude differences from EF and FB battery scores | 35 |
List of Figures

Figure 1. P3b waveform from an oddball paradigm 7
Figure 2. P3b memory-updating process 8
Figure 3. Example trial of ERP paradigm 13
Figure 4. Sample-by-sample t-tests results for adult participants 26
Figure 5. Topography of adult P3b effect 27
Figure 6. Sample-by-sample t-tests results for child participants 28
Figure 7. Topography of child P3b effect 29
Figure 8. Mean P3b amplitudes for adults and children for each condition 33
Figure 9. Exemplary waveforms showing the interaction effect 34
Figure 10. Scatter plots of P3b amplitude differences against EF and FB battery scores 37
## List of Abbreviations and Symbols

<table>
<thead>
<tr>
<th>Term</th>
<th>Abbreviation or Symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Theoretical or Technical Terms</strong></td>
<td></td>
</tr>
<tr>
<td>Electroencephalography / Electroencephalogram</td>
<td>EEG</td>
</tr>
<tr>
<td>Event-Related Potential</td>
<td>ERP</td>
</tr>
<tr>
<td>Executive Functioning</td>
<td>EF</td>
</tr>
<tr>
<td>False Belief</td>
<td>FB</td>
</tr>
<tr>
<td>(Right) Temporal Parietal Junction</td>
<td>(r)TPJ</td>
</tr>
<tr>
<td>Theory of Mind</td>
<td>TOM</td>
</tr>
<tr>
<td><strong>Statistical and Measurement Terms</strong></td>
<td></td>
</tr>
<tr>
<td>Analysis of Variance</td>
<td>ANOVA</td>
</tr>
<tr>
<td>Degrees of Freedom</td>
<td>df</td>
</tr>
<tr>
<td>Hertz</td>
<td>Hz</td>
</tr>
<tr>
<td>Independent Components Analysis</td>
<td>ICA</td>
</tr>
<tr>
<td>Mean</td>
<td>M</td>
</tr>
<tr>
<td>Microvolts</td>
<td>μ</td>
</tr>
<tr>
<td>Milliseconds</td>
<td>ms</td>
</tr>
<tr>
<td>Pearson Correlation Coefficient</td>
<td>r</td>
</tr>
<tr>
<td>Probability of Outcome</td>
<td>p</td>
</tr>
<tr>
<td>Sample Size</td>
<td>n</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>SD</td>
</tr>
<tr>
<td>Standard Error</td>
<td>SE</td>
</tr>
</tbody>
</table>
### Glossary

<table>
<thead>
<tr>
<th><strong>Concept or Conceptual</strong></th>
<th>“Concepts are the units of thought, the constituents of beliefs, and theories…” (Carey, 2009, p. 5)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Executive Functioning</strong></td>
<td>Multifaceted set of abilities necessary to purposefully execute goal-directed activity, such as planning, keeping goals in mind, the capacity to inhibit prepotent responses, and the ability to self-monitor (Anderson, 1998; Henry &amp; Bettenay, 2010).</td>
</tr>
<tr>
<td><strong>Explicit</strong></td>
<td>Purposeful and declarative. Direct task demands are <em>explicit</em> and typically require <em>explicit</em> (purposeful, declarative) responses.</td>
</tr>
<tr>
<td><strong>Implicit</strong></td>
<td>Automatic, non-declarative, and spontaneous. Indirect task demands are <em>implicit</em> and typically elicit <em>implicit</em> (automatic, non-declarative, and spontaneous) responses.</td>
</tr>
<tr>
<td><strong>Representational Theory of Mind</strong></td>
<td>The everyday understanding that others’ behaviour is at least partly motivated by internal representational mental states, such as beliefs, desires, or intentions (Wellman, 1990). A representational understanding of mental states is based on and constrained by individual experience and consists of the understanding that one’s mental state can be discordant with another’s or with a known reality.</td>
</tr>
</tbody>
</table>
Chapter 1: Introduction

"Everything you know about another person’s thoughts or feelings is inferred from his behaviour. Knowledge of behavior is factual. Knowledge of mental processes is theoretical, or inferential" (Hebb, 1966)

Theory of mind refers to our everyday understanding that human behaviour is at least partly motivated by internal representational mental states, such as beliefs, desires, and intentions (Wellman, 1990). Aspects of theory of mind begin to emerge at a young age; children follow others’ gaze as early as 12 months (Johnson, Slaughter, & Carey, 1998), and engage in pretend play and reasoning about diverse desires at around 2 years of age (Lillard, 1993, Rapeholi & Gopnik, 1997, Wellman). These abilities demonstrate an interest in others’ perspectives and mental states and an ability to engage in the mental role-playing of imagined characters. More importantly for the purpose of this study, these abilities constitute the precursors of representational theory of mind, hallmarked by the understanding of false beliefs.

False belief understanding – the understanding that others can hold beliefs that contradict the actual state of the world – is often tested with the unexpected-transfer task, originally developed by Wimmer and Perner (1983). In this task, children are introduced to a character who places an object in one of two possible locations and then leaves. While the character is absent, the object is moved to a second possible location. The character then returns, and the child is asked where the character is going to look for the object. Individuals who understand false belief will pass this task by correctly predicting that the character will look where she last left it, and not where the object truly is. Children begin to pass this false belief task between the ages of 3 to 5 years, with most young 3 year olds failing the task, and most 5 year olds passing the task (Amsterlaw & Wellman, 2006; Wellman, Cross, & Watson, 2001; Wellman & Liu, 2004). Children who
pass the standard false belief task are said to have a *representational theory of mind*, which is based on and constrained by individual experience and consists of the understanding that one’s mental state can be discordant with another’s or with a known reality. The ability to explicitly reason about another’s false belief requires a representational understanding, which is why I focus on belief understanding as an index of a mature theory of mind.

The unexpected transfer task, which is the standard assessment of false belief understanding, is dichotomous in that children can either pass or fail. The binary nature of this task lends itself to stage-like and unitary portrayals of the development of theory of mind (e.g., Wellman & Liu’s 2004 five-stage scaled theory of mind developmental progression), in which children suddenly shift from failing to understanding false beliefs. This stage-like framework implies that false belief understanding – the hallmark of representational theory of mind – exists only once children begin passing the standard false belief tasks.

In contrast to the standard developmental account detailed above – whereby young children first understand desires around the age of 2 years and only come to understand false belief around the ages of 4 or 5 years – there is evidence that young children can succeed on non-verbal, implicit assessments of false belief much before they are able to explicitly express that understanding. For example, Onishi and Baillargeon (2005) measured 15-month olds’ behaviour on a non-verbal false belief task and found that the toddlers’ looking times fit with an understanding of others’ mental states, including an understanding of beliefs. Onishi and Baillargeon’s (2005) used a violation of expectancy paradigm that consisted of three phases. First, infants watched familiarization
trials in which an actor puts an object in one of two possible locations and then reaches back into that location twice. Second, there was a belief induction trial in which the object stays in the same location, or is displaced to the second location as the actor either watches or does not watch. The first two belief induction trials (where the object does not move or the actor sees its displacement) would lead the actor to hold a true belief of the object’s location, whereas the last type of belief induction trial (the actor does not see the object’s displacement) would lead the actor to hold a false belief of the object’s location. The final phase of the experiment was a single test trial in which the actor reaches into one of the two locations. In half of the trials, this search was consistent with the actor’s belief, in the other half of the trials, the actor searched in the box that was not consistent with her belief. Onishi and Baillargeon (2005) found that infants looked longer when the actor’s search contradicted her belief, supporting the authors’ conclusion that infants predict others’ actions using belief reasoning. Other studies using violation of expectancy paradigms produced similar results, with 15 month infants and 2.5 year old toddlers looking longer at the belief-incongruent outcomes (Trauble, Marinovic, & Pauen, 2010; Scott, He, Baillargeon, & Cummins, 2012, respectively).

Studies using anticipatory looking paradigms to assess infants and children’s understanding of beliefs have yielded converging evidence (Surian & Geraci, 2012; Garnham & Ruffman, 2001). For example, Garnham and Ruffman (2001) had 2.5 year olds watch a scene in which one character places a ball in one of three locations, and then leaves. During his absence, a second character moves the ball to another of the three locations, and then leaves. The first character returns and children are prompted by the
narration “I wonder where he’s going to look”. Children looked more towards the first location (the belief-congruent location) than towards the other two locations (the empty location and the actual location of the ball – both of which are belief-incongruent locations). Garnham and Ruffman (2001) suggest that this pattern of results reflects 2.5 year olds’ sensitivity to others’ beliefs.

There are two main interpretations of these violation-of-expectancy and anticipatory-looking findings. First, one could argue that that infants and young children rely on a representational understanding of beliefs in order to succeed on these implicit assessments of false belief understanding. Baillargeon, Scott, and He (2010) take this stance: “Contrary to traditional claims, the ability to attribute false beliefs to others is already present by the second year of life”. By this argument, young infants and toddlers possess an underlying representational competence yet are unable to translate that competence into correct performance on the standard false belief task as a result of limiting factors such as language abilities, memory capacity, and executive functioning skills. In other words, performance limitations gate accurate performance on assessments such as the standard false belief task (Carlson & Moses, 2001).

An alternative explanation is that infants rely on non-conceptual cognitive tools in order to succeed on the implicit assessments of false belief (e.g., Apperly & Butterfill, 2009; Ruffman, 2014; Sabbagh, Benson, & Kuhlmeier, 2013). There are many opinions on the exact nature of these non-representational understandings. For example, Apperly and Butterfill (2009) suggest that infants possess an inflexible belief-tracking system,

---

1 Garnham and Ruffman’s (2001) paradigm actually includes a few more experimental features to control for a number of alternative hypotheses about how children solve the task. Here, I discuss only the features essential to my introduction. I will return to this study in the discussion section.
analogous to infants’ number-tracking system, which allows infants to succeed on implicit assessments of false belief understanding. Sabbagh, Benson, and Kuhlmeier (2013) argue that young infants pass implicit assessments of false belief understanding by engaging behavioural typologies that were built up by a combination of statistical learning mechanisms and episodic encoding of the features of others’ actions. Similarly, Ruffman (2014) states that infants’ early ability to pass implicit assessments of false belief results from a combination of statistical learning mechanisms and initial biases towards human faces and motion. Despite the differences in these explanations, there is an important unifying feature: all propose that early theory of mind performance is supported by a fundamentally different type of cognitive system than later theory of mind performance.

For a number of reasons, analyzing the available eye-gaze evidence alone does not provide a definitive answer to the question of whether young children’s performance on implicit false belief tasks is a result of representational understanding of beliefs or rather some other, perhaps statistical-based system. One reason is that the explicitness or implicitness of the task does not necessarily indicate the type of understanding (representational or non-representational) used to solve the task. It is possible that representational understandings can underpin correct performance on an indirect, or implicit, measure of belief understanding (Diennes & Perner, 1999). Thus, the use of an implicit false belief task affords little insight into the nature of the assessed understanding. Second, past studies have found that there is dissociation between event-related potential (ERP) and looking-time measures in infants on simple visual tasks (Nelson & Collins, 1991; Nikkel & Karrer, 1994). For example, infants’ event-related
potential responses distinguished between familiar visual events presented frequently versus infrequently as well as infrequent familiar and novel visual events, whereas their looking-time measures did not differ across conditions (Nelson & Collins, 1991). These results suggest that looking-time or eye-gaze measures do not provide a full picture of the cognitive tools that infants and young children use to process unexpected (or possibly statistically infrequent) outcomes. Indeed, Nikkel and Karrer (1991) suggest that both ERP and behavioural measures (such as looking time) should be used to understand early cognitive development.

Of particular interest in my study is the question of whether infants and young children rely on a representational understanding of beliefs to succeed on implicit false belief tasks. Because eye-gaze measures cannot definitively answer this question, I turned to a particular ERP, which has been related to contextual and semantic understandings: the P3b response. The P3b response is a positive component of the event-related potential (ERP) that occurs approximately 300-500 ms post-stimulus over parietal regions (Polich, 2007). This ERP response has been associated with memory-updating processes in response to unexpected events (Conroy & Polich, 2007; Knight, 1996; Overbeek et al., 2005; Polich, 2007; Polich & Criando, 2006). The task used to establish this finding is an oddball task in which participants perceive a stream of stimuli (such as simple images or tones), most of which are ‘background’ or distractor stimuli, some of which are targets, and some of which are novel stimuli (Figure 1). Participants are instructed to respond to the targets (e.g., by pushing a button), which differ only slightly from the distractors. In contrast, participants do not respond to the novel stimuli, which differ substantially from the distractors. In this task, the novel and target stimuli both elicit P3a, a reflection of
attention orientation preceding the P3b, but only the target tends to elicit a strong P3b. In patients with hippocampal damage, the P3a is elicited in response to targets, but not the P3b, suggesting a mechanism that involves memory-updating in response to relevant unexpected events (see Figure 2; Knight, 1996; Polich & Criando, 2006).

Figure 1. Event-related potential measured at Pz in response to a visual oddball paradigm with difficult visual discriminations (Comchero & Polich, 1999). The ‘target’ condition in Comchero and Polich’s (1999) study corresponds to the ‘surprise’ P3b response. Reprinted from Clinical Neurophysiology, 110, Comchero, M. D., & Polich, J., P3a and P3b from typical auditory and visual stimuli, 24-30, Copyright (1999), with permission from Elsevier.
Although the oddball paradigm uses relatively simple stimuli (e.g., tones and simple images), the P3b response has been observed in more cognitively complex tasks – such as responding to others acting in accordance with or in contradiction to their beliefs (Meinhardt, Sodian, Thoermer, Dohnel, & Sommer, 2011). Meinhardt and colleagues measured adults and 6 year old’s P3b responses to seeing a character act in either a belief-congruent or belief–incongruent way, and found that participants showed a stronger P3b response to seeing others act unexpectedly. These findings indicate that P3b is sensitive to conceptual understandings – it is not simply sensitive to statistical regularities, but seems to sensitive to violations of conceptual, perhaps representational frameworks. Furthermore, Meinhardt and colleagues (2011) suggested that the P3b may reflect the updating of “a mental model of the environment”, and that the amplitude of the P3b reflects the unexpectedness of the eliciting event (Meinhardt et al., 2011), providing
a way to measure how surprising an unexpected event is – or alternatively, how certain someone is in the theory that formed the contradicted prediction.

An added advantage of using the P3b response to investigate young children’s versus adults’ neural responses to belief-based events is that the P3b will allow us to not only index young children’s non-verbal expectations of others’ behaviour, but will also allow us to measure the strength of those expectations. As mentioned, the amplitude of the P3b response likely reflects the degree of unexpectedness of an outcome (Meinhardt et al., 2011) – that is, how surprised someone is by the outcome. It can be assumed that the degree of certainty in a prediction or expectation is proportionate to the degree of surprise at seeing that theory contradicted. So, by measuring P3b, it is possible to index how certain young children and adults are in their predictions of others’ behaviour. Measuring certain young children versus adults certainty in their predictions of others’ behaviour will help to clarify the differences that exist in their understandings of others’ mental states. Ruffman, Garnham, Import, and Connolly (2001) attempted to address this very point by having children use tokens to indicate how certain they were of their answers on a false belief task – the more tokens they bet, the more certain they were. However, this method is flawed because the ability to assess one’s own uncertainty is a metacognitive ability that is likely associated with representational theory of mind (Bartsch and Estes, 1996; Lockl & Schneider, 2007), introducing a critical confound. By leveraging the P3b response to a non-verbal false belief task, it will be possible to index young children’s certainty in their expectations without running into this confound.

Through the measurement of the P3b response, it was possible to investigate two main questions. First, I used the P3b response to address the question of whether
representational theory of mind understanding is indeed present at a young age, but perhaps is masked by performance limitations. To address this question, I investigated whether young children’s P3b responses to a non-verbal false belief match that of adults, whom most would argue have a representational understanding of beliefs. Second, I investigated whether young children’s behavioural performance on an executive functioning and a false belief battery predicted their P3b responses.

To address these questions, I created an unexpected-transfer with ERP recordings time-locked to target events. Children aged 3.5 to 4.5 (who typically would fail the standard false belief task) and adults watched 48 simple cartoons with the same basic structure, which can be described as a version of the standard false belief task. In the cartoons, a character placed an object in one of two containers, and then left the room. A second character then switched the location of the object and then left the room (now the first character had a false belief of the object’s location). The first character then appeared on the screen, leaving with one of the two containers – one of the containers was consistent with the character’s false belief of the object’s location (the belief-congruent condition), and the other was inconsistent with that false belief (the belief-incongruent condition). Event-related potentials were measured in response to the first character appearing with one of the two containers. The cartoons were structured such that half of the time, the first character left with the belief-congruent box, and the other half of the time, the first character left with the belief-incongruent box.

In addition to the ERP paradigm, child participants also completed two behavioural batteries. The first battery assessed executive functioning, which is a multifaceted set of abilities necessary to purposefully execute goal-directed activity such
as planning, keeping goals in mind, the capacity to inhibit prepotent responses, and the
ability to self monitor (Anderson, 1998; Henry & Bettenay, 2010). Because behavioral
measures of executive functioning and theory of mind are correlated (Hughes, 1998),
understanding the relationship between executive functioning and neural responses
reflecting belief understanding was of interest in the current study. The second battery
assessed children’s performance on explicit measures of false belief understanding,
affording a developmental context to the neural results, and making the findings of the
current study comparable to those of past studies.

In line with past research, I hypothesized that adults would show a greater P3b
response to seeing characters act in contradiction to their false beliefs (the belief-
incongruent condition) compared to seeing characters act according to their false beliefs
(the belief-congruent condition). For young children, the two positions on the nature of
their false belief understanding lead to two distinct predictions. If young children do
indeed possess a representational theory of mind, as posited by Baillargeon, Scott, and He
(2010), among others, then their pattern of P3b responses should not differ from that of
adults. That is, young children should show a greater P3b response in the belief-
incongruent condition compared to the belief-congruent condition. In contrast, if young
children are not solving the false belief task in the same way as adults, but rather are
using some non-representational mechanism, as suggested by Ruffman (2014) and
Sabbagh, Benson, and Kuhlmeier (2013), among others, then their P3b responses should
be the opposite of adults. That is, young children should show a greater P3b response in
the belief-congruent condition compared to the belief-incongruent condition. To
summarize, children showing a similar pattern to adults would support the conclusion
that young children pass implicit false belief tasks as a result of a representational understanding of beliefs. In contrast, a significant interaction will support the conclusion that young children pass implicit false belief tasks using a different, non-representational understanding of beliefs.

Chapter 2: Method

Participants

Thirty-eight children (19 girls, 23 boys) between the ages of 3.5 and 5.5 years of age and twenty-five adults (25 women, 2 men) between the ages of 18 and 19 were recruited to participate in this study. To recruit children for this study, I contacted families listed in the Psychology Department’s Developmental Database by phone or email. Adults were recruited through a first-year psychology course subject pool, which is why there were mostly women participants – predominantly women are enrolled in the first-year psychology course. To thank families for their participation, we gave them a $10 gift card to Chapters. Adults received a course credit towards their first year psychology course. (Please see results section for participant attrition and cases not included in the final analyses).

Materials

Stimuli. I created a multi-trial theory of mind task, composed of 48 brief cartoons, with one target event for ERP collection per animation. All of the cartoons followed the same structure, which was based on Scott’s (2013) paradigm and which can be described as an unexpected-transfer task. In each cartoon, there were two containers (one gift bag and one gift box), one object, and two characters (Figure 3). Participants were told that the characters in the cartoons are always going to parties and are always in
a rush to get off to the party, so they must grab the gift, jump on their bike, and head to the party. In the cartoons, the first character placed the object in one of the two containers; live narration accompanied this event (e.g., “Chloe wraps the cookies in the pink box!”). The character then left the room, and both containers were left closed (ex. narration “Chloe then leaves the room”). The second character then entered the room and switched the object from its original location to the other container (ex. narration “Max moves the cookies and wraps them in the green bag instead”). The second character left, and both containers were again left closed. After the second character left, the scene with the closed containers was accompanied by a live narration such as “It’s time to go to the party! Let’s see which one Chloe grabs!”. The last scene of the animation depicted the first character biking away with one of the two containers in the back basket of the bike. Half of the time, the container held the object in contradiction to his or her false belief (the belief-incongruent condition), and the rest of the time, the character left with the empty container in accordance with his or her false belief (belief-congruent condition). The container in the basket was brighter than the rest of the scene in order to emphasize and draw attention to the character’s choice. This last scene was time-locked for ERP collection.

*Figure 3. Example trial from the ERP paradigm.*
False Belief Battery. Children completed four standard false belief tasks that assessed the understanding that beliefs can differ from reality, which are each described in the following paragraphs (see appendix A for complete scripts). The false belief battery had good internal consistency in my sample, Cronbach’s α = .78. The inter-item correlations (i.e. the correlations among subscale scores) ranged from $r = .331$ to $r = .672$. The results of a principle axis factor analysis indicated that all four items loaded well on one dimension, with an unrotated eigenvalue of 2.41. The item loadings on the single factor ranged from .692 to .855.

(1) Contents Change task (Gopnik & Astington, 1988): Children were shown a bandaids box and were asked “What do you think is inside?”. Most children replied “bandaids”, but were shown that there were actually crayons inside the box. The box was then closed again and children were asked “Before we opened this box, before you looked inside, what did you think it had inside, bandaids or crayons?”. A similar question was asked with respect to another character: “Sally has never seen inside this box before, what does she think is inside: bandaids or crayons?”. The last question was a control question: “What’s really inside, bandaids or crayons?”. The correct pattern of responses to these three questions was: bandaids, bandaids, and crayons, respectively.

(2) Location Change task (Wimmer & Perner, 1983): The standard unexpected-transfer false belief task. Children were introduced to two characters, Doggy and Birdy, who were playing with a ball. Doggy leaves for lunch, and so puts the ball under a hat. After Doggy leaves, Birdy takes the ball to play some more with it, but then also wants to leave for lunch, and so places the ball in a basket. Doggy returns and children are asked the following questions: “Where does Doggy think the ball is?” (belief test question), and
“Where is the ball really?” (reality control question). The correct pattern of responses was: hat, and basket, respectively.

**(3) Appearance-Reality (Flavell, Green, & Flavell, 1986):** Children are shown a sponge that is painted to look like a rock, and were permitted to feel the ‘sponge-rock’ to learn that it really is a sponge, despite its’ appearance. Children were then asked “What does this look like, a rock or a sponge?” and “What is it really, a rock or a sponge?”. The correct pattern of responses was: rock, and sponge, respectively.

**(4) Deceptive Pointing task (Carlson, Moses, & Hix, 1998):** Children were introduced to a toy car and two boxes, and were shown that they could put the car in one of the boxes and then point to the box that has the car in it, so we know where it is. A research assistant, who was present during this familiarization, then left the room so that we could put a car in one of the boxes and see if she could find it upon her return. While she was gone, I said to the child, “Why don’t we play a trick on [research assistant’s name] so that she won’t find the car? We can play a trick by pointing so that she won’t find it”. When the research assistant returned, she asked “Where is the car?”. Correct responses, demonstrating an understanding of how to deceive, involved pointing to the empty box. The same was procedure was repeated a second time, with me leaving the room instead.

**Executive Functioning Battery.** Children completed five standard executive functioning tasks that assessed children’s ability to keep multiple rules in mind and to inhibit prepotent responses (please see appendix B for complete scripts). The executive functioning battery had acceptable internal consistency in my sample, Cronbach’s $\alpha = .60$. The inter-item correlations ranged from $r = -.034$ to $r = .475$. The results of a
principle axis factor analysis indicated that there were two factors (Initial eigenvalues of 1.994 and 1.059), with some items cross-loading. The pointing stroop, dimensional change card sort task, and the day night task all loaded on the first factor. The whisper task and the bear dragon task cross-loaded on both factors. The item loadings on factor 1 ranged from .405 to .865; the item loadings on factor 2 ranged from .013 to .693. Although the factor analysis suggested a 2-factor solution, I decided to run my analyses using the full battery in order to make my findings comparable to the many past studies that have used this standard battery of tasks.

(1) Bear/Dragon task (Reed, Pien, & Rothbart, 1984). Children were introduced to two puppets: a nice bear and a mean dragon. Children were told that we should listen to the bear, but not do what the dragon tells us to do. The bear and dragon then alternately gave 5 commands each such as “touch your nose” and “stick out your tongue”. Children were scored based on how well they responded to the bear’s commands and how well they inhibited responses to the dragon.

(2) Pointing Stroop Grass/Snow task (Carlson & Moses, 2001): In this task, children had to point to a green card in response to hearing the command “snow”, and to a white card in response to hearing the command “grass”. There were 16 trials in total, and children’s score consisted of how many correct responses they gave (in the form of pointing).

(3) Day/Night task (Gerstadt, Hong, & Diamond, 1994): In this task, children had to say “day” in response to seeing a black card with the moon and stars, and “night” in response to seeing a white card with a sun on it. There were 16 trials in total, and children’s scores consisted of how many correct responses they gave.
(4) *Dimensional Change Card Sort task (Zelazo, 2006):* Children were first told that we were going to play a shape game, where all the rabbit cards should go into the bin with the blue rabbit label, and all of the boat cards should go into the bin with the red boat label. The cards to be sorted consisted of rabbits and boats, some of which were red and some of which were blue. After children had sorted all of the 5 cards, I introduced a rule change: we were then going to play the color game, in which all the blue things go into the bin with the blue rabbit label, and all of the red things go into the bin with the red boat label. Children’s performance was scored based on their success sorting the second set of 5 cards, after the rule change was introduced.

(5) *Whisper task (Kochanska, Murray, Jacques, Koenig, & Vandegeest, 1996):* Children were told that we were going to play a whisper game, where it was necessary to whisper the whole time. They were show cards of some well-known and less popular cartoon characters, and were asked to whisper the names of the characters. Children were scored on each trial based on how well they adhered to the whisper rule – the lowest score was given if the child yelled the character’s name, and the highest score was given if they whispered.

**EEG/ERP Recording and Data Processing**

Net. The electroencephalogram (EEG) was recorded using a Hydrocel Geodesic Sensor Net that consisted of 128 Ag/AgCl electrodes that are connected in an elastic geodesic tension structure that distributed the electrodes evenly when the net was fitted onto the head. EEG activity in all channels was recorded in reference to the vertex electrode (Cz), sampled at 250 Hz, and digitally filtered between .01 and 200 Hz (time
constant = 1s). EEG was recorded continuously throughout the 48 cartoons. Impedences were brought to below 40Ω before beginning the trials.

Procedure

Arrival. For child participants, upon arrival in the laboratory, a research assistant spent some time playing and coloring with the child, ensuring that the child was comfortable. While the child was getting familiarized, I explained the ERP recording process and general procedures to the parents. After the parents read through the information form and signed the consent form, and after the child had given verbal assent, testing began with the ERP procedure.

For adult participants, I explained the ERP recording process, allowed them time to read over the information form, and obtained their informed consent.

ERP Collection. For child participants, during the EEG recording process, children sat at a table in a booster seat. Parents were in the testing room, near the child, at all times. Before the net was placed on the child’s head, it was soaked in a mild electrolyte solution (1.5 teaspoons of potassium chloride and a drop of baby shampoo in 1L of filtered water). While the net was soaking, we measured and lightly marked (with a soft, washable crayon) the vertex of the child’s head; this mark was used to guide net placement. When the child was ready, I shook excess water off of the net and then stretched the net over the child’s head such that all of the electrodes made contact with the scalp evenly. Once the net was in place, I worked to make sure that all electrodes were positioned properly, while a research assistant helped the child with fruit snacks and a juice box. Once the net was properly on and the impedences were below the desired 40Ω, the experimental task (the 48 brief cartoons) started, accompanied by live narration.
Eight ‘sticker breaks’ were randomly interspersed throughout the task in order to afford a break and to keep children interested in the task.

For adult participants, generally the procedure above applied with a few exceptions: adults sat in an adjustable computer chair rather than a booster, and much to their dismay, adults did not receive snacks or stickers.

**False Belief and Executive Functioning Batteries.** Only child participants completed the false belief and executive functioning batteries, as it was assumed that adults would perform at ceiling on all tasks. After children watched all of the ERP task cartoons, and the net was removed, children completed the false belief battery and the executive functioning tasks. The tasks were presented in alternating order: a false belief task, then an executive functioning task, followed by a false belief task and so on.

**Data Cleaning and Processing**

Raw EEG signals were first filtered using a .3 – 20 Hz bandpass digital filter. The time-locked epochs were extracted the segments were 1000 ms long following the target event, with a 100 ms baseline. Target events consisted of either the belief-congruent, or the belief-incongruent outcome. Bad channels were detected by inspection and were then interpolated using a spherical spline in the Net Station software. Bad channels were those whose amplitudes varied largely and incongruously with contiguous channels, likely a result of a bad wire connection or poor contact with the participant’s scalp. Next, paroxysmal trials (caused by participants’ large movements) were deleted, and then the data were processed using an independent components analysis (ICA; Makeig, Bell, Jung, & Sejnowski, 1996) algorithm. Artifactual components – including ocular and muscle artifacts – were removed and the data were reconstituted.
A second round of bad channel detection and interpolation followed using the following algorithm: channels that had a greater range than 200 μV in more than 4 trials were considered bad channels for adults and were deleted and interpolated. For children, the criteria were more liberal: channels that had a greater range than 300 μV in more than 4 trials were considered bad channels. The criteria for the child data were more liberal because children’s data tend to be much nosier due to more movement, and applying the same criteria as adults would cause the loss of too many trials to proceed with analyses. Alternatively, applying the child criteria to the adult data would lead to the unnecessary inclusion of noisy trials in the adult data. Essentially, the criteria were set to balance the two main objectives of data processing: (1) minimize artifacts, and (2) maximize available trials for analyses.

After channel interpolation, bad epochs were deleted using a similar method: epochs were considered bad if more than 4 channels in an epoch had a range greater than 150 μV for adults. For children, the criteria more liberal: epochs were considered bad if more than 4 channels had a range greater than 350 μV. Finally, the data were base-line corrected. The same justification outlined above applies to this difference in criteria.

This stream of processing was based on: (1) Net Station waveform tools guidelines (Electrical Geodesics, 2003), (2) eeglab resources (Delorme & Makeig, 2004), and (3) past experience.

Chapter 3: Results

Participant Attrition and Cases not Included in Analyses

Child Participants. In total, 38 eligible children participated in this study. However, a number of participants were not included in the final sample of children. The
breakdown detailing the reasons for which children were excluded from analyses is presented in Table 1.

<table>
<thead>
<tr>
<th>Age Range</th>
<th>Total Participated</th>
<th>Dropped</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.5-4.0</td>
<td>16</td>
<td>2</td>
</tr>
<tr>
<td>4.0-4.5</td>
<td>14</td>
<td>2</td>
</tr>
<tr>
<td>4.5-5.0</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>5.0-5.5</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>38</td>
<td>8</td>
</tr>
</tbody>
</table>

Table 1. The number of child participants in each age range that were dropped for various reasons from the analyses. Four additional children were recruited to participate in the study, but were not eligible for various reasons: the research assistant was ill ($n = 2$), the participants showed up too late to run the whole study ($n = 1$), and one participant was the twin of another participant ($n = 1$). During data processing, participants were dropped if they did not have enough artifact-free trials (see methods for criteria).

To ensure that our task – specifically sitting through a 30 to 40 minute EEG session – was not selecting the children with high executive functioning, I compared children’s EF and FB battery scores on the basis of whether or not they completed the EEG task. The EF and FB battery scores were computed by first standardizing the scores on each of the tasks, and then averaging the standardized scores for each battery.

Children who completed the EEG task did not significantly differ on age, EF, or FB score from those who did not complete the EEG task (Table 2).

<table>
<thead>
<tr>
<th>Completed EEG</th>
<th>Did not Complete EEG</th>
<th>$t$ (df)</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>4.18 (0.53)</td>
<td>4.19 (0.60)</td>
<td>0.059 (36)</td>
</tr>
<tr>
<td>FB Battery</td>
<td>0.12 (0.75)</td>
<td>0.056 (1.00)</td>
<td>0.18 (31)</td>
</tr>
<tr>
<td>EF Battery</td>
<td>0.046 (0.62)</td>
<td>-0.15 (0.82)</td>
<td>0.67 (31)</td>
</tr>
</tbody>
</table>

Table 2. Results from independent samples t-tests comparing the age, FB, and EF standard battery scores of those who completed versus those who did not complete the EEG task.
**Adult Participants.** Of the 25 adults that participated in this study, 2 participants were excluded from analyses. There was a computer error for one participant that caused data loss, and one participant was dropped during data processing because there were not enough artifact-free trials. Two additional participants were not included in the number of eligible adult participants because one had just-dyed hair (which would leech into the EEG net), and there was not a large enough EEG hat for another participant.

**Behavioural Results**

Only child participants completed the behavioural tasks, which consisted of a false belief (FB) battery of four tasks and an executive functioning (EF) battery of five tasks.

**Inter-rater reliability.** Two independent coders scored participants’ behavioural performance on each of the nine tasks. Overall, there was a high concordance between the two coders: their scores for the FB battery were highly significantly correlated, $r(35) = .952, p < .001$, as were their scores for the EF battery, $r(35) = .961, p < .001$. These reliability scores were computed using all of the participants with available behavioural data ($n = 37$). Details for scoring the tasks can be found in appendices A and B.

**Subscale Raw Scores.** The raw mean raw scores for all participants with behavioural data and for the participants included in analyses are displayed in Table 3.
All Participants  
(n=37)  
M (SD)  
Sample for Analyses  
(n=19)  
M (SD)  

<table>
<thead>
<tr>
<th>False Belief Battery</th>
<th>All Participants</th>
<th>Sample for Analyses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deceptive Pointing (/2)</td>
<td>0.98 (0.94)</td>
<td>1.21 (0.98)</td>
</tr>
<tr>
<td>Location Change (/1)</td>
<td>0.30 (0.44)</td>
<td>0.26 (0.45)</td>
</tr>
<tr>
<td>Contents Change (/2)</td>
<td>0.89 (0.87)</td>
<td>0.95 (0.90)</td>
</tr>
<tr>
<td>Appearance/Reality (/2)</td>
<td>1.31 (0.75)</td>
<td>1.45 (0.72)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Executive Functioning</th>
<th>All Participants</th>
<th>Sample for Analyses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Day/Night (/18)</td>
<td>8.69 (5.53)</td>
<td>9.50 (5.85)</td>
</tr>
<tr>
<td>Bear/Dragon (/34)</td>
<td>30.19 (6.15)</td>
<td>31.92 (4.34)</td>
</tr>
<tr>
<td>Pointing Stroop (/18)</td>
<td>11.97 (4.97)</td>
<td>12.13 (4.78)</td>
</tr>
<tr>
<td>Card Sort (/5)</td>
<td>4.26 (1.05)</td>
<td>4.33 (1.02)</td>
</tr>
<tr>
<td>Whisper Task (/2)</td>
<td>1.84 (0.30)</td>
<td>1.83 (0.32)</td>
</tr>
</tbody>
</table>

Table 3. Mean scores and standard deviations for each subscale. The maximum score is indicated by the name of each subscale – for example, the maximum score on deceptive pointing is 2.

Correlations. As past research has found that age, executive functioning, and false belief performance are significantly positively correlated in preschoolers (Hughes, 1998), I ran pearson correlations among these three variables to assess whether these relationships existed in my data. Indeed, age was significantly positively correlated with both FB and EF scores, and FB and EF scores were significantly positively correlated with each other in the full sample of participants (see Table 4). With the exception of the significant relationship between age and EF score, these correlations were also present in the subset of participants used for the analyses (Table 4). The sample used in analyses had less variability in age compared to the full sample of participants, which may explain why one of the correlates of age was not significant in the subsample.
Table 4. Pearson correlations among age, FB, and EF Battery scores

<table>
<thead>
<tr>
<th></th>
<th>All Participants</th>
<th>Sample for Analyses</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( r )</td>
<td>( r )</td>
</tr>
<tr>
<td>Age with FB Battery</td>
<td>.637( ** )</td>
<td>.606*</td>
</tr>
<tr>
<td>Age with EF Battery</td>
<td>.561( ** )</td>
<td>.326</td>
</tr>
<tr>
<td>FB Battery with EF Battery</td>
<td>.676( ** )</td>
<td>.713( ** )</td>
</tr>
</tbody>
</table>

** \( p \leq .001 \)
* \( p \leq .01 \)

ERP Findings

To investigate participants’ P3b responses to seeing others act according to, or in contradiction to their false beliefs, I first looked at the overall topography of the expectation differences in adults and children. To do this, I ran paired sample-by-sample t-tests on each electrode for both age groups. I used the results from these t-tests to help characterize the P3b effect and to guide focused analyses.

Although I investigated the overall pattern of effects and then focused in on a particular area, it is important to note that these investigations were theoretically and empirically based. As mentioned in the introduction, the P3b event-related potential occurs in parietal regions approximately 300-500 ms post-stimulus. I used these criteria, along with the morphology of a P300 wave from oddball paradigms to characterize the P3b effect in adults and children. In past studies, researchers using sparse electrode arrays have used a single electrode, typically Pz (electrode 62), which is on the midline, over the parietal region, to index the P3b response. However, taking this limited perspective does not take into account both the detailed topography of the P3b response, which can differ with age and stimulus type (Polich 2007; Comerchero & Polich, 1999). Of particular importance in the present context is that the P3b response in this social-cognitive context might be right-lateralized, given that belief reasoning is related to rTPJ activity (e.g.,
Sabbagh, Bowman, Evraire, & Ito, 2009; Saxe, 2010; Scholz, Triantafyllou, Whitfield, Brown, & Saxe, 2009; Perner, Aichorn, Kronbichlen, Staffen & Ladurner, 2006). For this reason, I characterized the P3b effect across groups of electrodes rather than a single, pre-determined electrode.

Sample-by-Sample Paired-Samples t-tests. To characterize the P3b effect in both adult and children, I conducted sample-by-sample paired samples t-tests for each electrode, with the following criteria to reduce the risk of false positives. First, I considered only those electrodes in the parietal region, within the time window of 300 to 500 ms. A significant P3b effect was defined as at least five significant continuous samples on at least 3 contiguous electrodes.

Adult Topography. Using the criteria described above, I found a clear P3b expectation effect starting around 400 ms and continuing until approximately 480 ms in parietal electrodes (Figure 4). As can be seen in Figure 5, adults tend to show higher P3b responses to seeing others act in accordance with their false beliefs in comparison to seeing others act in contradiction to their false beliefs. Also in Figure 5, it can be seen that the findings are slightly right-lateralized. Based on the latency and topography of the effect, as well as the wave morphology (Figure 5), these differences likely reflect a P3b effect.
Figure 4. The results from the sample-by-sample paired samples t-tests for each electrode. The highlighted time window from approximately 300-500 ms was the time window of interest for the P3b wave. The plot shows the $p$ values (according to the color scale at the right) for each time sample for each electrode. The white box highlights the observed P3b effects (early and late).
Figure 5. The topographical plot in the center indicates which channels showed a P3b expectation effect based on the sample-by-sample paired samples t-tests. Sample waveforms from the circled electrodes are displayed on the sides; the dashed line denotes trials in which the protagonists acted in contradiction to their beliefs, whereas the solid line denotes trials in which the protagonists acted in accordance with their beliefs. On these waveform plots, time (in ms) is along the horizontal axis, amplitude (in μV) is along the vertical axis, and grey shading indicates $p < .05$ for the pairwise comparison, where darker shading reflects lower $p$ values. The yellow highlighted regions from approximately 300-500 ms indicate the time window of interest for the P3b wave.
Child Topography. Using the previously described criteria, I found a P3b expectation starting around 400 ms and continuing until approximately 530 ms (see Figure 6). Although P3b tends to occur 300-500 ms post-stimulus, it is not unreasonable that children’s P3b responses were protracted because past research shows that children’s P3b effects tend to be later than that of adults (Polich, Ladish, & Burns, 1990). As can be seen in Figure 7, children tend to show higher P3b responses to seeing others act in contradiction to their false beliefs in comparison to seeing others act in accordance with their false beliefs. Figure 7 also demonstrates the right-lateralization of the effect. Like the adult ERPs, based on the latency and topography of this effect, as well as the wave morphology, these differences likely reflect a P3b effect.

![Figure 6](image_url)

*Figure 6.* The results from the sample-by-sample paired samples t-tests for each electrode for children participants aged 3.5-4.5 years (*n* = 17). The highlighted time window from approximately 300-500 ms was the time window of interest for the P3b wave. The plot shows the *p* values (according to the color scale at the right) for each time sample for each electrode. The white box highlights the observed P3b effects.
Figure 7. The topographical plot in the center indicates which channels showed a P3b expectation effect based on the sample-by-sample paired samples t-tests. Sample waveforms from the circled electrodes are displayed on the sides; the dashed line denotes trials in which the protagonists acted in contradiction to their beliefs, whereas the solid line denotes trials in which the protagonists acted in accordance with their beliefs. On these waveform plots, time (in ms) is along the horizontal axis, amplitude (in μV) is along the vertical axis, and shading indicates $p < .05$ for the pairwise comparison, where darker shading reflects lower $p$ values. The yellow highlighted regions from approximately 300-500 ms indicate the time window of interest for the P3b wave.
Comparing Adults and Children’s P3b Responses. When comparing adult and child ERPs, it is important to bear in mind that the adult and child brains do not necessarily overlap in terms of the topography of their responses. Comparing matching electrodes (i.e. comparing channel 62 in adults to channel 62 in children) may not be informative, because characteristics of P3b response is affected by age (Polich et al., 1990). Therefore, I used the wave morphology and sample-by-sample paired samples t-tests to guide my analyses. It was my aim to quantitatively capture the observed differences in the response patterns of children and adults.

In the adult data, the early P3b effect is present on the following channels: 62, 72, 77, 78, and 79, with a time window of approximately 392 to 472 ms post-stimulus. For this comparison, I averaged across the five electrodes for the time window 392:472 ms to produce an average for the false belief congruent and the false belief incongruent conditions. I averaged these electrodes, rather than analyze the electrodes individually, because the recorded amplitude at contiguous electrodes is not independent. Supporting this assertion is the fact that the electrodes’ averaged amplitudes in the 392:472 ms time window are all highly correlated in both conditions (see Table 5). Thus, testing each of these electrodes separately would result in analyses assessing overlapping variance, possibly leading to an inflated risk of false positives.
Table 5. Pearson correlations among averaged amplitudes of contiguous channels. The time window for amplitude averaging was 392 to 472 ms post-stimulus onset. The df for all correlations is 21.

**p<.001

In the child data, the early P3b effect is present on the following channels: 84, 89, and 90, with a time window of approximately 412 to 524 ms. Although channel 91 did show significant differences around the P3b peak, these differences mostly preceded the peak, a pattern which is different from that of channels 84, 89, and 90. This time window is slightly later than that of the adult time window, consistent with the fact that ERP responses tend to occur later in children compared to adults (Polich et al., 1990). Like the adult data, I computed the average amplitude across the electrodes for each condition in the time window of 412:524 ms post-stimulus. These three channels were highly correlated in both conditions (Table 6).
To further characterize children and adults’ P3b responses to the belief -
incongruent versus -congruent conditions differed, I ran a 2 (outcome) X 2 (age category) repeated-measures ANOVA. The dependent variable of this analysis was the average amplitude of the P3b channels for each condition. There was a significant outcome X age interaction, $F(1,38) = 13.02, p = .001$, partial eta squared = .244, observed power = .940. This interaction exists because adults showed a significantly higher P3b response in the belief-incongruent condition ($M = 3.69, SE = 0.596$) compared to the belief-congruent condition ($M = 2.64, SE = 0.518$) , $t(22) = 2.835, p = .01$, whereas children showed a significantly higher P3b response in the belief-congruent condition ($M = 18.80, SE = 3.14$) compared to the belief-incongruent condition ($M = 14.10, SE = 2.90$), $t(15) = 2.621, p = .019$. Essentially, children and adults’ P3b responses are flipped relative to one another (Figure 8). For added interpretation, an exemplar ERP from each age group is displayed for comparison in Figure 9.

<table>
<thead>
<tr>
<th></th>
<th>Belief-Incongruent</th>
<th>84 with 89</th>
<th>r</th>
<th>84 with 90</th>
<th>r</th>
<th>89 with 90</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td>Belief-Congruent</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>84 with 89</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>84 with 90</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>89 with 90</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Table 6. Pearson correlations among averaged amplitudes of contiguous channels. The time window for amplitude averaging was 392 to 472 ms post-stimulus onset. The df for all correlations is 15.*

**p<.001
*p<.05
Qualified by the interaction, there was a main effect of outcome, $F(1,38) = 5.22, p = .028$, partial eta squared = .121, observed power = .605, such that the FP had a significantly higher average amplitude ($M = 10.72, SE = 1.38$) compared to that of the FC condition ($M = 8.90, SE = 1.29$). This effect was likely driven by the fact that children’s ERP responses in both categories ($M = 16.54, SE = 1.93$) were significantly higher than that of adults ($M = 3.16, SE = 1.66$), $F(1,38) = 27.30, p < .001$, partial eta squared = .418, observed power = .999, in conjunction with the fact that children’s ERP responses were highest in the FP condition.

*Figure 8.* Mean amplitudes in the FC and FP condition for adults and children. The means for adults were calculated by averaging channels 62, 72, 77, 78, and 79 in the time window of 392:472 ms post-stimulus. The means for children were calculated by averaging channels 84, 89, and 90 in the time window of 412:524 ms post-stimulus. Error bars represent standard error.  

*p* $< .05$, based on a paired-samples t-test.
Figure 9. Waveforms to demonstrate the cross-over interaction effect. The solid line denotes trials in which the protagonists acted in contradiction to their beliefs, whereas the dashed line denotes trials in which the protagonists acted in accordance with their beliefs. Shading indicates \( p < .05 \) for the pairwise comparison, where darker shading reflects lower \( p \) values.

**Predictors of the Early P3b Effect in Young Children.** To investigate whether behavioural measures of EF and FB predicted the difference in ERP amplitude, controlling for age, I ran a linear regression. The outcome variable was the averaged amplitude difference from 412 ms to 524 ms between the belief- incongruent and congruent conditions for the following for channels: 84, 89, and 90. The outcome variable was computed by subtracting the belief congruent condition from the belief-incongruent condition, meaning that higher positive scores corresponding to a more ‘adult-like’ pattern of P3b responses. The time window was based on the average latencies of the P3b effect at each channel. I chose these three channels because they are topographically grouped and share a similar wave morphology and pattern of effects. I averaged these electrodes, rather than analyze the electrodes individually, because the recorded amplitude at contiguous electrodes is not independent. Supporting this assertion is the fact that the amplitude differences at these three electrodes were moderately to highly correlated (see Table 7).
Overall, the linear combination of EF and FB accounted for a significant proportion of the variance in amplitude difference, adjusted $R^2 = .339$, $F_{\text{change}}(2, 13) = 4.85$, $p = .027$. Of the individual predictors, EF battery score significantly contributed to the prediction of amplitude difference: greater EF scores predicted greater amplitude differences (Table 8). The FB battery score also significantly contributed to the prediction of amplitude difference: greater FB scores predicted lower amplitude differences (Table 8).

<table>
<thead>
<tr>
<th>Predictor</th>
<th>$\beta$</th>
<th>$t$ $(df)$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>EF Battery Score*</td>
<td>.860</td>
<td>3.01 (13)</td>
<td>.010</td>
</tr>
<tr>
<td>FB Battery Score*</td>
<td>-.747</td>
<td>2.62 (13)</td>
<td>.021</td>
</tr>
</tbody>
</table>

Table 8. Regression coefficients for the prediction of amplitude difference.

To aide the interpretation of these results, I have included scatter plots of the relationship between amplitude differences and each battery scores (Figure 10). On these plots, the red dashed line indicates the threshold between positive and negative amplitude differences; values above this line reflect cases that showed a higher P3b response to the belief-incongruent condition, and values below this line reflect cases that showed a higher P3b response to the belief-congruent condition.

The first thing to note is that there are only a few ($n = 4$) cases that have an amplitude difference above zero, meaning P3b responses were higher in the belief-incongruent condition. Of these four, three of the cases are hovering near the zero line, and one is an outlier with a high positive amplitude difference (at 9.02 $\mu$V).
As seen in the figure, higher EF scores are related to higher amplitude differences, with most cases still below the zero line (Figure 10). In contrast, higher FB scores are related to lower amplitude differences; it is important to note that the 0 line falls within the bounds of the 95% confidence interval, indicating a null zero-order relationship between FB and amplitude difference.
Figure 10. Scatter plots between the amplitude difference and battery scores. Along the vertical axis is the difference belief-incongruent minus belief-congruent amplitude difference. Along the horizontal axis is the average battery score (EF on top and FB on bottom). The red line dashed line indicates the ‘no difference’ threshold for the amplitude difference. Straight lines represent lines of best fit. Curved lines represent 95% confidence interval of the trend line. The zero order correlation between EF and the amplitude difference was not significant, $r(14) = .354, p = .179$, nor was the zero order correlation between FB and the amplitude difference, $r(14) = -.165, p = .542$. 
Chapter 4: Discussion

In this study, I investigated young children’s and adults’ P3b responses to seeing others act either in accordance with (belief-congruent condition), or in contradiction to their beliefs (belief-incongruent). The purpose of this study was to assess the neural mechanisms underlying young children and adults’ performance on an implicit false belief task with the aim of addressing the question of whether young children’s early success on implicit false belief tasks is supported by a representational understanding of beliefs. I found a significant interaction between age group and outcome such that adults’ P3b responses were higher in the belief-incongruent condition compared to the belief-congruent condition, whereas children’s P3b responses were higher in the belief-congruent condition than in the belief-incongruent condition. These findings, which contradict young children’s behavioural performance on implicit false belief tasks, support the conclusion that young children do not rely on the same representational understanding of beliefs that is present in older children and adults.

Performance on Implicit False Belief Tasks and the 2-System Approach

Past studies using anticipatory gaze and violation of expectancy paradigms have found that infants and young children, who do not pass the explicit unexpected-transfer task, show eye gaze patterns on implicit false belief tasks that are congruent with an understanding of beliefs (e.g., Onishi & Baillargeon, 2005; Surian & Geraci, 2012; Scott, He, Baillargeon & Cummins, 2012). That is, children look towards the belief-congruent location in anticipation of a character’s search, and will look longer when characters search in the unexpected location (i.e., the belief-incongruent location). In contrast to these findings, young children’s P3b responses in this study did not accord with an
understanding of false beliefs. Children’s P3b responses were higher, indicating more surprise, when characters searched in the belief-congruent location than when they searched in the belief-incongruent location. Understanding the import of this difference is a main focus of this discussion.

Some researchers have suggested that eye-gaze may index predictive sensitivities that are implicit and statistically-based (Dienes & Perner, 1999; Ruffman, 2014; Sabbagh, Benson, & Kuhlmeier, 2013). This suggestion is interesting in light of the fact that many of the eye-gaze studies employ ‘familiarization trials’, in which participants watch two or three instances of how things work or how characters behave in the paradigm, before they complete the ‘test trials’ (e.g., Onishi & Baillargeon, 2005, Surian & Geraci, 2012, and Trauble, Marinovic, & Pauen, 2010 used familiarization trials). Of course, none of these studies trained participants on the false-belief situation before the test trial, but it is noteworthy that these familiarization trials are not part of the standard, explicit false belief task (nor were they part of my study). In the standard false belief task, participants are expected to pass using their existent representational understanding of beliefs, without the preceding in-task contextual cues. In eye-gaze studies, familiarization trials may act as implicit prompts, activating the statistical-based frameworks that may underlie infants and young children’s successful performance on eye-gaze tasks. If the tasks indexed a representational understanding of beliefs, such familiarization trials would not seem necessary – adults and older children automatically tend to explain behaviour in terms of mental states, with minimal context or cues to do so. For example, Heider and Simmel (1944) showed participants a video of geometric shapes moving around a scene, and participants almost always described the actions mentalistically, attributing agency to the
shapes. Adults and older children with developed representational theories of mind do not need prompts to interpret actions mentallyistically – it is something that we just do.

The standard, explicit false belief task and implicit assessments also differ in that children can show dissociable performance in the same study using these two measures. That is, children’s eye-gaze patterns may reflect an understanding of beliefs, while their explicit responses reflect a lack of understanding of beliefs. Garnham and Ruffman (2001) measured children’s (~2-4 years old) anticipatory looks in response to the prompt “I wonder where he’s going to look”, and then measured their verbal responses to the question “Where is Sam going to look first?”. Whereas only 19% of participants answered the verbal question correctly, 72% looked towards the correct (belief-congruent) location (Garnham & Ruffman, 2001). These results suggest that the representational understanding indexed by the standard false belief task is not the same as the understanding underlying success on implicit assessments of false belief. Fitting in with this story, the child participants in the present study (aged 3.5 – 4.5 years), who were performing at approximately chance on average on the false belief tasks, showed P3b responses that reflected a lack of understanding of others’ beliefs. In contrast, adults, whom we assumed possessed a developed representational understanding of others beliefs, showed P3b responses that reflected an understanding of others beliefs. To tie these three measures together, it would interesting to measure children’s eye-gaze, P3b responses, and verbal response on the same false belief task. Based on the evidence from my study and on past research, I would expect that the P3b response would be linked to the verbal response, and that these two responses would be dissociable from eye-gaze. In other words, I propose that verbal and P3b responses on the false belief task are a
function of the same underlying understanding, whereas eye-gaze likely indexes a different understanding.

Taken together with past results, such as the dissociation of eye-gaze responses from verbal responses, the present finding that the P3b response reflects verbal responses on the false belief task support a 2-system interpretation of theory of mind development. Adults and older children who pass the explicit false belief task likely rely on the later-developing, representational understanding of others’ mental states, whereby they are able to reason explicitly about others’ mental representations. This representational understanding is likely related to the P3b responses of adults in this study. Infants and young children who fail the explicit false belief task but pass the implicit assessment likely rely on an early-developing system based on statistical or pattern recognition, behavioural typologies, or perhaps a belief-tracking system (Apperly & Butterfill, 2009; Dienes & Perner, 1999; Ruffman, 2014; Sabbagh, Benson, & Kuhlmeier, 2013). In this study, this early understanding was not reflected in young children’s P3b responses, suggesting that the early-developing theory of mind system is supported by a different neural or conceptual mechanism. Further investigations into the neural mechanisms supporting early sensitivity to beliefs would help support this claim, and would help clarify the nature of infants and young children’s early theory of mind understandings.

**Predictors of P3b Amplitude in Young Children**

Although the restricted age range was a limitation in my study, I did find a significant regression model predicting the difference in P3b amplitude (belief-incongruent minus belief-congruent P3b amplitude) in young children. A negative difference score would be consistent with the P3b pattern observed in young children,
where the P3b response was higher in the belief-congruent condition. In contrast, a positive difference would be consistent with the adult pattern, where the P3b response was higher in the belief-incongruent condition. In the regression, both the false belief and executive functioning battery scores were significant predictors of the amplitude difference.

Specifically, higher EF scores predicted a greater amplitude difference, with most amplitude differences still below the 0-difference line. In other words, although most children showed higher P3b responses in the belief-congruent condition (leading to a negative difference), there was an upward trend towards a difference of zero as EF scores increased. Having approximately equal P3b responses to both conditions would indicate uncertainty in so far as the child has no strong prediction that the character will look in one box over the other. This regression finding suggests that young children with lower EF are quite certain in their predictions that others will look for objects where those objects really are, whereas young children with higher EF are not so sure what to expect. It would be interesting to investigate a broader age range – in this study, I was only able to assess the regression results in children aged 3.5 – 4.5 years of age, however I expect that this trend would continue as children developed. That is, older children with higher EF scores will likely show increasing P3b amplitudes, getting into positive differences, reflecting an adult-like pattern.

The finding that higher EF scores predict increases in P3b difference fits with past evidence that aspects of EF performance are related to false belief understanding (Benson, Sabbagh, Carlson, & Zelazo, 2013; Carlson, Moses, & Claxton, 2004; Drayton, Turley-Ames, & Guajardo, 2011), and more specifically, that EF predicts improvements
in false-belief performance (Benson et al., 2013). Benson and colleagues (2013) suggest that EF is related to theory of mind development because improvements in EF allow children to benefit from relevant learning experiences. The role of EF in allowing children to benefit from experience fits squarely with the context-updating hypothesis of P3b, stating that P3b reflects concept updating in response to unexpected task-relevant events. Clearly, the relationship between EF and the P3b response needs further investigation, but the link found in this study is consistent with past research and makes sense within the theoretical framework of this study.

In contrast to the EF regression results, higher FB scores predicted a lower, more negative, amplitude difference in the regression model. This result is little more difficult to interpret, but looking at a scatter plot of the zero-order relationship between FB and the amplitude difference is helpful. Whereas the EF scatter plot shows a clear trend of increasing EF related to increasing amplitude difference, the FB scatter plot is less clear. The zero-line (no relationship) is within the 95% confidence bounds of the relationship between FB and the amplitude difference, indicating that, when taken alone, there is not a significant relationship between FB batter scores and amplitude difference. However, in the regression model, FB is a significant predictor of amplitude difference, suggesting that perhaps it accounts for some variance in amplitude difference that EF does not account for. The relationships among the battery scores and the P3b amplitude differences would likely be clarified by running a similar study with a broader age range.

**Implications for the Development of Theory of Mind**

Although not addressed directly by my study, the main findings have implications for understanding the development of theory of mind. In my study, young children
showed a higher P3b response to seeing others act in accordance with their false beliefs, a finding that is consistent with the view that young children expect others to act on reality rather than on beliefs. However, by age 6, children show a P3b response pattern that is more similar to that of adults (Meinhardt et al., 2011), whereby they show higher P3b responses to seeing others act in contradiction to their beliefs. Presumably, sometime between the ages of 4.5 and 6 years, the early pattern of P3b responses changes to resemble that of the adult pattern. What is the trajectory of this shift in neural responses?

I propose that the ERP amplitudes shift gradually to resemble the pattern present in older children and adults, and more specifically that P3b responses reflect a mechanism for changes in theory of mind (as also suggested by Meinhardt et al., 2011). This proposition is consistent with the view that representational theory of mind develops gradually, and is consistent my finding that higher EF predicts higher P3b amplitude differences.

Despite the fact that many behavioural assessments of theory of mind understanding imply a stage-like developmental progression, there is much evidence that theory of mind development is gradual. First, children do not simultaneously pass all forms of the false belief task at once, even if those tasks are equally difficult (e.g., Jenkins & Astington, 1995). This finding suggests that there is a period in which children can reason correctly about false beliefs in some contexts but not in other contexts.

Second, if 3 year olds that initially fail the standard unexpected-transfer task are given more opportunities to encode the details of the false belief task (e.g., spending more time attending to each event, guided by an experimenter), some are then able to pass the false belief task (Lewis, Freeman, Hagestad, & Douglas, 1994). This finding suggests that children who fail false belief tasks differ in their levels of belief understanding; for some
children, false belief understanding is within their zone of proximal development (Vygotsky), but for others, false belief understanding is yet unattainable. Finally, versions of the false belief task that use continuous, rather than dichotomous, response formats – such as response time and continuous rather than discrete search locations – have found that the tendency to choose the ‘belief-incongruent’ location decreases continuously with time (Apperly, Warren, Andrews, Grant, & Todd, 2011; Atance, Bernstein, & Meltzoff, 2010; Sommerville, Bernstein, & Meltzoff, 2013). These findings collectively undermine the simple pass/fail dichotomy of false belief tasks.

Fitting with the evidence that theory of mind develops gradually, the P3b response is a strong candidate for the mechanism underlying representational theory of mind development. From this study, we find that young children’s P3b responses are the flip of adults’ responses, but by age 6, their P3b responses resemble that of adults (Meinhardt et al., 2011). The P3b response has been linked to activity in the temporal parietal junction (TPJ; Polich & Criando, 2006), and activity in the right TPJ (rTPJ) has been related to selectively reasoning about others’ beliefs (Sabbagh, Bowman, Evraire, & Ito, 2009; Saxe, 2010; Scholz, Triantafyllou, Whitfield, Brown, & Saxe, 2009; Perner, Aichorn, Kronbichlen, Staffen & Ladurner, 2006). Finally, the P3b is thought to reflect a theory-updating mechanism in which unexpected events are integrated into and update the current theory (Meinhardt et al., 2011; Polich, 2007). Taken together, these findings provide an impetus for continuing to investigate the role of P3b, and the mechanism that it reflects, in theory of mind development. Although my project did not have a large enough age range to address this gradual development question, applying similar methods to a broader age range would start to get at this interesting developmental question.
**P3b Topography and the Neural Correlates of Theory of Mind**

The topography of the P3b responses in children and adults are consistent with past research on the neural correlates of theory of mind. Specifically, the P3b effect observed in both age groups occurred in mostly right-lateralized electrodes. The P3b response has been associated with activity in the TPJ (Polich & Criando, 2006), and the rTPJ has been linked to selectively reasoning about others’ beliefs (e.g., Sabbagh, Bowman, Evraire, & Ito, 2009; Saxe, 2010; Scholz, Triantafyllou, Whitfield, Brown, & Saxe, 2009; Perner, Aichorn, Kronbichlen, Staffen & Ladurner, 2006). What is interesting in this study, however, is that young children’s P3b responses are right lateralized over parietal regions, yet their P3b responses do not reflect an understanding of false belief. Even though young children do not seem to be reasoning at a representational level about beliefs, their ‘surprise’ responses to seeing others act in contradiction to their expectations occurs in generally the same regions as adults’ surprise responses. It is possible that the rTPJ may be involved more generally in processing expectations and outcomes related to others’ behaviour. The basis on which these predictions are formed may develop, yet where they are processed may by-and-large remain the same. Past research suggests that the development of the rTPJ is associated with improvements in theory of mind (Sabbagh, Bowman, Evraire, & Ito, 2009), and perhaps this development is somehow linked to shifts in the frameworks that are used to form predictions of others’ behaviour. The P3b response may play a role in this shift, as there is evidence that it is involved in updating current theories or understandings in response to contradictory evidence (Polich, 2007).

**Limitations**
There are two main limitations of this study that should be addressed in future investigations of belief understanding. First, as mentioned earlier, the restricted age range of this study (3.5-4.5 years) did not afford insight into the developmental trends of P3b as it relates to behavioural measures of executive functioning and false belief understanding. Extending the age range to include children who are passing the standard verbal false belief task would clarify these developmental trends. Second, although there is evidence that P3b reflects conceptual understandings, more research is needed on the neural mechanism reflected by this potential and specifically how it relates to cognition. For example, some research suggests the P3b response is related to activity in the temporal parietal junction, whereas other studies have linked it to activity in the locus-coeruleus (Polich & Criado, 2006). More research on the neural basis of P3b and how it is linked to cognition will help to clarify its role in the development of representational belief understanding.

Conclusions

The results of this study contribute to our understanding of the nature of young children’s belief understanding. Young children’s P3b responses to seeing others act based on beliefs suggests that they do not yet have a representational understanding of theory of mind. However, past research has found that infants and toddlers are able to succeed on non-verbal assessments of false belief understanding, which in conjunction with our findings supports the idea that there are two systems underlying theory of mind performance. The early developing system, which relies on implicit understandings and accounts for infants and toddlers’ performance, does not likely rely on the mechanism reflected by P3b responses. In contrast, the later, representational understanding of theory
of mind does seem to be underpinned by the mechanism reflected by P3b. These results point to a possible mechanism for change of theory of mind, whereby P3b responses reflects theory updating in response to unexpected events, consistent with my finding that higher EF predicted a more adult-like P3b pattern. Applying methods similar to this study to investigate a broader age range would help to clarify the role of the mechanism reflected by P3b in theory of mind development.
References


Surian, L., & Geraci, A. (2012). Where will the triangle look for it? Attributing false beliefs to a geometric shape at 17 months. *British Journal of Developmental Psychology, 30*, 30-44.


Appendix A: False Belief Battery Scripts and Scoring

1. Contents Change False Belief

Look what I have here. What do you think is in this box? **Band aids** other:____

It's a Bandaid box, isn’t it? Go ahead and look inside. (open box for S)

Hey! Those are Crayons! I wonder what Crayons are doing in a Bandaid box.

(close box)

When you first saw this box, before you opened it, did you think it had Band aids or Crayons inside?

**Self Response:** Band aids(1) Crayons(0)

Here’s Ernie. He has never looked inside this box before. What does he think is inside, Band aids or Crayons?

**Other Response:** Band aids(1) Crayons(0)

What is really inside, Band aids or Crayons?

**Control:** Band aids(0) Crayons(1)

OK. Let's show him. (Ernie says, "hey, what are those Crayons doing in there?")

(close box, put Ernie away)

OK. You're doing a great job!

Note: Child must get the control question correct to receive credit on any false belief question.

SELF SCORE: 0 1

OTHER SCORE: 0 1

**Total Unexpected Contents Score:** ____ out of ____.

2. Location Change False Belief

(place blue box (B) on child's left and red box (A) on child's right)

Now let me show you what else I have. Here are two puppets. This one is Ernie (make Ernie say hello). And this one is Burt (make Burt say hello).
Ernie and Burt are playing with this ball. (make puppets play together).
Burt is gonna go eat lunch now. He'll put the ball in here (place ball in blue box; Burt hidden away).
Ernie wants to play some more. (Ernie takes out ball, plays).
Now Ernie is gonna go eat lunch. He'll put the ball in this box (places car in red box; Ernie hidden away).
Here comes Burt again. He wants to play with the ball some more.

Where does Burt think the ball is?
B-Blue(1)   A-Red(0)   Other ______
(prevent S from opening box)

Where is it really?
B-Blue(0)   A-Red(1)   Other ______

You can go ahead and show Burt where the ball is.
Ok! That was fun!

Overall Score:    Fail(0)    Pass(1)
(S must answer both questions correctly to pass)

3. Appearance-Reality

Sponge/Rock
Look what I have here. What does this look like? rock  other ______
It looks like a rock, doesn't it? But really and truly it's a sponge. See (E squeezes sponge and offers to let S touch it).
When you look at this right now, does it look like a sponge or like a rock?

sponge(0)  rock(1)

OK.
What is this really and truly, a sponge or a rock?

sponge(1)  rock(0)

OK.

Sponge/Rock Score:   Fail(0)    Pass(1)
(S must answer both questions correctly to pass)

Red/Black Castle
Now look at this. This is a red castle. But look, when I put this cover on it, it looks black, see?
Really and truly it's red (uncover) but we can make it look black (cover).
When you look at the castle right now, does it look red or black?

red(0)  black(1)

OK. What color is the castle really and truly, red or black?
red(1)  black(0)

Castle Score: Fail(0)  Pass(1)

OK. You’re doing a great job!  

Total A/R Score: ___

4. Deceptive Pointing

Warm-up trial (pointing = location)
We’re gonna play a game with E2.  See this car?  And see these boxes over here?  
I’m gonna put the car inside one of the boxes like this (E1 puts car in gray box).  
I can point to the box with my finger like this (points) so we’ll know which box the 
car is in. 
Now you try.  Take the car out of the box.  Go ahead and put the car in that box (blue 
box).  Ok, now, point to that box so we’ll know where the car is.  
Good job!  See, now we can tell where the car is by pointing to it.

Test Trial # 1
Now let’s put a car in a box and see if E2 can find it.  E2 is going to leave and go in 
another room so he/she can’t see.  (E2 exits)  
Ok. Go ahead and put the car in one of the boxes.

Car placed in:  GRAY  BLUE

Hey, I have a great idea!  Let’s play a funny trick on E2.  Let’s play a trick so he/she 
can’t find the car.  Maybe we could trick him/her so he’ll/she’ll look in the wrong 
box.  OK? 
Now, remember, we’re gonna play a funny trick on E2.  We can play a trick by 
pointing so he/she won’t find the car.  
Here comes E2. (E2 returns and asks) Where is the car?

Prompt 1:  E2 is asking where the car is. Where do you want to point? 
Prompt 2: (if no response): Which box do you want to point to? 
Prompt 3: (if still no response): Do you want to point to this box (gray) or this box 
(blue)?

Child points to:  GRAY  BLUE

OK E2. You can look now.
(If child uses deception, E2 looks in the wrong box and E1 says)
Oh, he/she didn't find it! I guess we tricked him/her! We are so tricky. Go ahead and show him/her where it really is.
(If no deception, E2 looks in the correct location and E1 says)
Oh, he/she found it! I guess we didn't trick him/her that time.
(E1 says) That was fun. Now it's my turn to leave the room. (Exits)

Test Trial # 2
Now let's put a car in a box and see if E1 can find it. E1 is going to leave and go in another room so he/she can't see. (E1 exits)
Ok. Go ahead and put the car in one of the boxes.

Car placed in:  GRAY  BLUE

Hey, I have a great idea! Let's play a funny trick on E1. Let's play a trick so he/she can't find the car. Maybe we could trick him/her so he'll/she'll look in the wrong box. OK?
Now, remember, we're gonna play a funny trick on E1. We can play a trick by pointing so he/she won't find the car.
Here comes E1. (E1 returns and asks) Where is the car?

Prompt 1: E1 is asking where the car is. Where do you want to point?
Prompt 2: (if no response): Which box do you want to point to?
Prompt 3: (if still no response): Do you want to point to this box (gray) or this box (blue)?

Child points to:  GRAY  BLUE

OK E1. You can look now.
(If child uses deception, E1 looks in the wrong box and E2 says)
Oh, he/she didn't find it! I guess we tricked him/her! We are so tricky. Go ahead and show him/her where it really is.
(If no deception, E1 looks in the correct location and E2 says)
Oh, he/she found it! I guess we didn't trick him/her that time.
That was fun. We're all done with that game. You did a great job!
It's fun to play tricks sometimes, but we should always tell the truth, OK?

# Deceptive Trials: ____

% Deception: ____
(# deceptive trials /total trials)
Appendix B: Executive Functioning Battery Scripts and Scoring

1. Bear/Dragon Task

Warm-up
(make sure S can follow all of the directions by having S do the following; E models the movements during warm-up)
OK, I’m going to ask you do do some silly things before we start our next game.
Stick out your tongue. (E remember to model)

Touch your ears.
Touch your teeth.
Touch your eyes.
Clap your hands.
Touch your feet.
Touch your head.
Touch your tummy.
Touch your nose.
Wave your hand.

OK. Good job!

Instructions
(E takes out the bear and dragon puppets)
I have a game we can play with these puppets. This puppet (show the bear) is a nice bear. When he talks to us, we will do what he tells us to do. This puppet (show the dragon) is not very nice. This puppet is a dragon. So when he talks to us, we won’t listen to him. If he tells us to do something, we won’t do it.

Practice Trials

OK. Let’s practice one time.
BEAR: This is the good bear. He says, "Touch your nose." (in a mellow, nice voice)

<table>
<thead>
<tr>
<th>touch</th>
<th>no touch</th>
<th># tries</th>
</tr>
</thead>
</table>

(If S does not touch nose)
Remember, we listen to the nice bear and do what he says because that’s how we play the game.

(E repeats the command and may model the action until S succeeds)

(IF S touches nose)
That’s right! Now let’s practice with the naughty dragon. In this game, we won’t do what the dragon asks us to do because he’s not so nice.

DRAGON: Touch your tummy. (in a low, gruff voice)
touch no touch # tries ___

(If S touches stomach)
Remember, we are not going to listen to the mean dragon. We won’t do what he says because that’s how we play the game.

(Repeat until S gets it right, holding S’s hands down if necessary on the 6th try)
(When S succeeds at both practice trials)
Yeah! That was fun.

Rule Check

So, when the Bear tells you to do something, do you do it? Yes No
(correct if necessary) # tries ___

And when the Dragon tells you to do something, do you do it? Yes No
(correct if necessary) # tries ___

OK, let’s play some more.

Test Trials
(E does not model or give feedback on test trials) Coder 1 Coder 2

1. BEAR: Stick out your tongue. ___ ___ ___
2. DRAGON: Touch your ears. ___ ___ ___
3. BEAR: Touch your teeth. ___ ___ ___
4. DRAGON: Touch your eyes. ___ ___ ___
5. BEAR: Clap your hands. ___ ___ ___

(Reminder given regardless of performance)

Remember the way we play this game. We do what the bear tells us to do because he is nice but we won’t do what the dragon tells us to do because he isn't nice.

6. DRAGON: Touch your feet. ___ ___ ___
7. BEAR: Touch your head. ___ ___ ___
8. DRAGON: Touch your tummy. ___ ___ ___
9. BEAR: Touch your nose. ___ ___ ___
10. DRAGON: Wave your hand. ___ ___ ___

Coding Instructions

Bear trials: Bear Total (add trials 1, 3, 5, 7, and 9): ___
0 = failure to move
1 = a wrong movement
2 = a partial correct movement  
3 = a full correct movement  

Dragon trials:  
**Dragon Total** (add trials 2, 4, 6, 8, and 10): ____  
0 = a full commanded movement  
1 = a wrong movement  
2 = a partial commanded movement  
3 = no movement  

**Strategies used on Dragon trials:** (check all that apply, add comments below)  
____  Says “no”  
____  Shakes head  
____  Sits on hands  
____  Other, describe:  

---  

2. **Pointing Stroop**  
(Bring out a board that has a solid white card attached to the upper left corner, a solid green card attached to the upper right corner (both cards are 15 x 10 cm), and two fabric cut-outs shaped like a child’s hands centered below the cards.)  

We’re going to play a game with this board.  
Do you know what colour grass is? ________  

Do you know what colour snow is? ________  

(praise right answers; correct wrong answers)  

Well, we’re gonna play a silly game.  
In this game, when I say the word “grass”, I want you to point with your finger to the white card like this (E points to white card).  

Can you point to the white card?  
(Praise if pointed, prompt if not)  

Pointed  
Did not point  

When I say the word “snow”, I want you to point with your finger to the green card like this (E points to green card).  

Can you point to the green card?  
(Praise if pointed, prompt if not)  

Pointed  
Did not point  

Practice Trials
(say “Grass” and wait for response)
(if S hesitates) What card do you point to for this one?  **green(0) white(1)**  
#tries

(praise if correct, say “Snow” and wait for response)
(if S hesitates) What card do you point to for this one?  **green(1) white(0)**  
#tries

(praise if correct, go on to test trials)
(if incorrect, repeat both rules starting with the one S got wrong, and repeat practice)

**Test Trials**
(whenever S hesitates, ask "What card do you point to for this one?" but do not use words "white" or green"; do not give feedback on test trials)

<table>
<thead>
<tr>
<th></th>
<th>RT1</th>
<th>RT2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Grass</td>
<td>green(0) white(1)</td>
<td></td>
</tr>
<tr>
<td>2. Snow</td>
<td>green(0) white(1)</td>
<td></td>
</tr>
<tr>
<td>3. Snow</td>
<td>green(0) white(1)</td>
<td></td>
</tr>
<tr>
<td>4. Grass</td>
<td>green(0) white(1)</td>
<td></td>
</tr>
<tr>
<td>5. Snow</td>
<td>green(0) white(1)</td>
<td></td>
</tr>
<tr>
<td>6. Grass</td>
<td>green(0) white(1)</td>
<td></td>
</tr>
<tr>
<td>7. Grass</td>
<td>green(0) white(1)</td>
<td></td>
</tr>
<tr>
<td>8. Snow</td>
<td>green(0) white(1)</td>
<td></td>
</tr>
<tr>
<td>9. Snow</td>
<td>green(0) white(1)</td>
<td></td>
</tr>
<tr>
<td>10. Grass</td>
<td>green(0) white(1)</td>
<td></td>
</tr>
<tr>
<td>11. Snow</td>
<td>green(0) white(1)</td>
<td></td>
</tr>
<tr>
<td>12. Grass</td>
<td>green(0) white(1)</td>
<td></td>
</tr>
<tr>
<td>13. Grass</td>
<td>green(0) white(1)</td>
<td></td>
</tr>
<tr>
<td>14. Snow</td>
<td>green(0) white(1)</td>
<td></td>
</tr>
<tr>
<td>15. Grass</td>
<td>green(0) white(1)</td>
<td></td>
</tr>
<tr>
<td>16. Snow</td>
<td>green(0) white(1)</td>
<td></td>
</tr>
</tbody>
</table>

Total % Correct: ___  
% Agreement: ___  

% Correct Trials 1-4: ___  
Median RT (Coder 1): ___  

% Correct Trials 13-16: ___  
Median RT Trials 1-4: ___  
Median RT Trials 13-16: ___  

**3. Day/Night Task**
**Instructions**

We're gonna play a game with these cards.

Do you know when the sun comes out? In the day or in the night? __________

Do you know when the moon and stars come out? In the day or in the night? __________

(praise right answers; correct wrong answers)

Well, we're gonna play a silly game.

(show a black moon card)

In this game, when you see this card, I want you to say "day".

Can you say the word "day"? said not said

(praise if said, prompt if not said)

(remove card and show white sun card)

When you see this card, I want you to say "night".

Can you say the word "night"? said not said

(praise if said, prompt if not said, remove card)

**Practice Trials**

(show white sun card and wait for response)

(if S hesitates) What do you say for this one? day(0) night(1) #tries

(praise if correct, show black moon card)

(if S hesitates) What do you say for this one? day(1) night(0) #tries

(praise if correct, go on to test trials)

(if incorrect, repeat both rules starting with the one S got wrong, and repeat practice)

**Test Trials**

(whenever S hesitates, ask "What do you say for this one?" but do not use words "day" or night"; do not give feedback on test trials)

RT1 RT2

1. (white sun card)    day(0) night(1) ___ ___ ___

2. (black moon card)   day(1) night(0) ___ ___ ___
3. (black moon card) day(1) night(0) ____
4. (white sun card) day(0) night(1) ___ ___
5. (black moon card) day(1) night(0) ___
6. (white sun card) day(0) night(1) ___ ___
7. (white sun card) day(0) night(1) ___ ___
8. (black moon card) day(1) night(0) ___
9. (black moon card) day(1) night(0) ___
10. (white sun card) day(0) night(1) ___ ___
11. (black moon card) day(1) night(0) ___
12. (white sun card) day(0) night(1) ___ ___
13. (white sun card) day(0) night(1) ___ ___
14. (black moon card) day(1) night(0) ___
15. (white sun card) day(0) night(1) ___ ___
16. (black moon card) day(1) night(0) ___

Total % Correct: ___ % Agreement: ___

% Correct Trials 1-4: ___ Median RT (Coder 1): ___
% Correct Trials 13-16: ___ Median RT Trials 1-4: ___
Median RT Trials 13-16: ___

4. Card Sort

Instructions

(E places trays on table. Tray on left has a red rabbit card pasted on back wall. Tray on right has a blue boat card pasted on back wall.)
We are going to play a game. This is the SHAPE game. All the rabbits go in this box (pointing to left tray) and all the boats go in that box (pointing to right tray). We don't put any rabbits in that box. No way (shaking head). We put all the rabbits over here and only boats go over there. If it is a rabbit, then it goes here. If it is a boat, then it goes there. This is the SHAPE game.

OK. I'll go first. Rabbits go here. (E places a blue rabbit card in slot on left)
Boats go here. (E places a red boat card in slot on right)
Pre-switch Trials

OK. Now it's your turn.
1. If it is a rabbit, then put it here, but if it is a boat, put it there.
   Here is a blue rabbit. Where does this go? rabbit tray* boat tray ___

2. If it is a rabbit, then put it here, but if it is a boat, put it there.
   Here is a red boat. Where does this go? rabbit tray* boat tray* ___

3. If it is a rabbit, then put it here, but if it is a boat, put it there.
   Here is a blue boat. Where does this go? rabbit tray* boat tray* ___

4. If it is a rabbit, then put it here, but if it is a boat, put it there.
   Here is a red rabbit. Where does this go? rabbit tray* boat tray ___

5. If it is a rabbit, then put it here, but if it is a boat, put it there.
   Here is a blue rabbit. Where does this go? rabbit tray* boat tray ___

(When S is correct)
Yes, that's right.

(When S is wrong)
No, that's not right. Remember the rules. (proceed to next trial)

(S must correctly sort 5 consecutive cards according to shape before proceeding to post-switch trials; remove cards from tray if needed)

Total Number Pre-switch Trials: ___

Number of Pre-Switch Errors: ___

Post-switch Trials

Now we are going to switch. We are not going to play the shape game any more. We are going to play the COLOR game. When it is red, you have to put it in this box (indicating left tray), but whenever it is blue, then it goes in that box (indicating right tray). We don’t put red things in that box. No way (shaking head). We put red things over here and only when it’s blue does it go over there. If it’s blue, then it goes there. If it’s red, then it goes here.

(E does not give feedback on post-switch trials)

1. If it is red, then put it here, but if it is blue, put it there.
   Here is a red boat. Where does this go? red tray(1) blue tray(0) ___

OK. Let’s do another.

2. If it is red, then put it here, but if it is blue, put it there.
Here is a red rabbit. Where does this go? \(\text{red tray}(1)\) \(\text{blue tray}(0)\) ___
OK. Let’s do another.

3. If it is red, then put it here, but if it is blue, put it there.
Here is a blue boat. Where does this go? \(\text{red tray}(0)\) \(\text{blue tray}(1)\) ___
OK. Let’s do another.

4. If it is red, then put it here, but if it is blue, put it there.
Here is a red boat. Where does this go? \(\text{red tray}(1)\) \(\text{blue tray}(0)\) ___
OK. Let’s do another.

5. If it is red, then put it here, but if it is blue, put it there.
Here is a blue rabbit. Where does this go? \(\text{red tray}(0)\) \(\text{blue tray}(1)\) ___

Compatible Trial Total (2 and 3): ___
Incompatible Trial Total (1, 4, and 5): ___
Post-switch Total: ___

5. **Whisper Task**

Instructions

For this game I need to see if you can whisper.
(E now starts to WHISPER THROUGHOUT TASK)

Can you whisper your name really quietly? \(\text{whisper} \quad \text{no whisper}\)

# tries ___

(Praise if whisper)
(If S does not whisper)
Let’s try to talk very very quietly, just like I am. Can you whisper your name to me?
(Continue even if child never whispers)

**Test Trials**

OK. Let’s play this whisper game I know. I have some pictures of some cartoon
characters. Let’s see if you can tell me their names. Some of them I bet you’ll know
and some might be kind of hard and you won’t know them, and that’s OK.
Remember to whisper because that’s how we play the game.
(For each card ask, ”Do you know who this is?”)

<table>
<thead>
<tr>
<th>Know?</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y</td>
<td>N</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Know?</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y</td>
<td>N</td>
</tr>
</tbody>
</table>
2. Pocahontas  Y N Other __________
3. Huckle       Y N Other __________
4. Donald Duck  Y N Other __________
5. Elmer Fudd   Y N Other __________

You are doing really well. Remember that this is a whisper game where we try to whisper the whole time. (said regardless of performance)

6. Snow White   Y N Other __________
7. Petunia      Y N Other __________
8. Beast        Y N Other __________
9. Fat Albert   Y N Other __________
10. Mickey Mouse Y N Other __________

(If S simply nods and says yes, prompt with, "Can you tell me his/her name?")
(If S shakes head or says no, then say "That's OK. This is a hard one. His/her name is ____. Let's try a different one.")
(If S gives an incorrect name, say "OK" and move on; code as usual but note the answer given.

Coding Instructions

**Carlson Method:**
Score only the items that S knew (or at least gave an answer to). If S did not answer (said nothing) or said "I don't know" then do not code it.

0 = shout
1 = normal voice or mixed--S starts in one mode of voice and then changes to another mode such as in self-correction (shout to whisper) or in a loss of control (whisper to shout)
2 = whisper

**Total Score:** ____

**# S answered:** ____

**Mean Score (ours):** ____

total/#answered
Appendix C: Ethics Approval and Amendment

April 01, 2013

Ms. Samantha Drover
Master's Student
Department of Psychology
Queen's University
Kingston, ON K7L 3N6

GREB Ref #: GPSYC-604-13; Romeo # 6007817
Title: "GPSYC-604-13 Brain responses, Social Experiences, and Theory of Mind"

Dear Ms. Drover:

The General Research Ethics Board (GREB), by means of a delegated board review, has cleared your proposal entitled "GPSYC-604-13 Brain responses, Social Experiences, and Theory of Mind" for ethical compliance with the Tri-Council Guidelines (TCPS) and Queen's ethics policies. In accordance with the Tri-Council Guidelines (article D.1.6) and Senate Terms of Reference (article G), your project has been cleared for one year. At the end of each year, the GREB will ask if your project has been completed and if not, what changes have occurred or will occur in the next year.

You are reminded of your obligation to advise the GREB, with a copy to your unit REB, of any adverse event(s) that occur during this one year period (access this form at https://eservices.queensu.ca/romeo_researcher/ and click Events - GREB Adverse Event Report). An adverse event includes, but is not limited to, a complaint, a change or unexpected event that alters the level of risk for the researcher or participants or situation that requires a substantial change in approach to a participant(s). You are also advised that all adverse events must be reported to the GREB within 48 hours.

You are also reminded that all changes that might affect human participants must be cleared by the GREB. For example you must report changes to the level of risk, applicant characteristics, and implementation of new procedures. To make an amendment, access the application at https://eservices.queensu.ca/romeo_researcher/ and click Events - GREB Amendment to Approved Study Form. These changes will automatically be sent to the Ethics Coordinator, Gail Irving, at the Office of Research Services or irvingg@queensu.ca for further review and clearance by the GREB or GREB Chair.

On behalf of the General Research Ethics Board, I wish you continued success in your research.

Yours sincerely,

John D. Freeman, Ph.D.
Professor and Acting Chair
General Research Ethics Board

cc: Dr. Mark Sabbagh, Faculty Supervisor
    Dr. Rod Lindsay, Chair, Unit REB
    Marie Tooley, Dept. Admin.
January 14, 2014

Ms. Samantha Drover  
Master’s Student  
Department of Psychology  
Queen’s University  
Humphrey Hall  
62 Arch Street  
Kingston, ON, K7L 3N6

Dear Ms. Drover:

RE: Amendment for your study entitled: GSYC-604-13 Brain responses, Social Experiences, and Theory of Mind; ROMEO# 6007817

Thank you for submitting your amendment requesting the following changes:

1) To recruit adult participants over the age of 18 from the Queen’s student population through the Psychology Department Subject Pool, e-mails sent out to undergraduate students who have asked to be notified of studies, and posters put in campus buildings;

2) Recruitment email for Adult Participants; the Subject Pool notice and the poster will contain key points taken from the recruitment email. (Note: Please refer to our Participant Recruitment Procedures on our website: http://www.queensu.ca/ers/researchethics/GeneralREB/forms.html).

By this letter you have ethics clearance for these changes.

Good luck with your research.

Joan Stevenson, Ph.D.  
Chair  
General Research Ethics Board  

c: Dr. Mark Sabbagh, Supervisor