Kwantlen Psychology Student Journal
Issue 1, July 2019
Kwantlen Polytechnic University

What Eye Tracking Reveals in Implicit-Discrete Versus Explicit Continuous Theory-of-Mind Measures

Angela L. Giesbrecht[1]

Abstract

Adults can understand others' mental states (Theory of Mind, ToM), but their private knowledge tends to hinder this ability (ToM errors). Eye-tracking recorded where adult participants looked in two ToM tasks. Firstly, participants watched videos where characters held a false or true belief about an animal's location. This task was implicit because it did not solicit a response from participants. As predicted, participants looked longer and first looked where a character, with a true belief, would search for an object. When a character had a false belief, however, participants looked longer and first looked at incorrect response locations. Secondly, adults listened to stories where characters held a false or true belief about an object's location. Participants indicated where characters would search for an object (search) and the object's initial location (memory). This task was explicit because it solicited a response from participants. Contrary to predictions, participants made more ToM errors when answering the search question in the true belief condition. In comparison, participants made fewer ToM errors when answering the search question in the false belief condition and the memory question in both belief conditions. Methodological issues may account for this discrepancy. The study suggests that adults make ToM errors.

Keywords: theory-of-mind (ToM), implicit ToM, explicit ToM, eye tracking, anticipatory looking

---

[1] angela.giesbrecht95@gmail.com; Written for an Honour's Thesis (PSYC 5000/5010). Angela plans to pursue a master's degree in Counselling Psychology at Trinity Western University and then work as a counsellor.

Acknowledgements – This work was supported by the Student Led Research Grant from the Kwantlen Polytechnic University Office of Research and Scholarship (SR18-08) to the author. The author would like to thank Dr. Daniel M. Bernstein, Daniel Derksen, Eric Mah, Dr. Jocelyn Lymburner, Kristen Zeller, Dr. Jeffery Foster, Isu Cho, and Dr. Adam Cohen for providing helpful feedback, guidance, and useful materials.

What Eye Tracking Reveals in Implicit-Discrete Versus Explicit-Continuous Theory-of-Mind Measures

Imagine you put your paper coffee cup on your desk when you come to your office. Then you leave to retrieve your briefcase from your car. In your absence, your co-worker recycles your cup. When you come back to your office, you look for your cup on your desk. This should not surprise your co-worker because this is where you last saw your cup. Therefore, you still believe that your cup is there. This example illustrates that individuals attribute the intentions and beliefs of other individuals in daily social situations. People manipulate, explain and predict others' behaviour in their daily lives (Apperly, Riggs, Simpson, Chiavarino, & Samson, 2006). This requires the ability to understand others' mental states, including what they believe, desire, and intend; this ability is called Theory of Mind (ToM) (Premack & Woodruff, 1978; Wellman, Cross, & Watson, 2001).

When individuals use their ToM, or infer what others believe, they have to recognize that other individuals' mental states may or may not be consistent with reality. To understand how others' beliefs might affect their behaviour, people have to recognize that others' beliefs can be *correct* (i.e., true or consistent with actual events and behaviours) or *incorrect* (i.e., false or inconsistent with actual events or behaviours). Thus, people can have false beliefs or true beliefs (herein referred to as *FB* and *TB*, respectively). In particular, children's ability to reason about another person's belief has been researched, especially when the other person has a false belief (Wellman et al., 2001).

Studying False Belief in Children

False belief has commonly been studied using the *change-of-location* task. In this task, participants hear a narrative about two characters, Sally and Anne, based on Wimmer and Perner's (1983) paradigm. In Baron-Cohen, Leslie, and Frith's (1985) adaptation, Sally puts an object (e.g., a marble) in a basket. When Sally leaves, a secondary character, Anne, moves the object to a box. When Sally returns, participants are asked where Sally will look for the object as well as where the object actually is and was initially; thus, participants are asked *belief*, *reality* and *memory* questions, respectively. Participants pass the task if they consider Sally's FB that the object is where she left it, and correctly predict that she will look in the first location. Participants also have to give correct answers to the reality and memory questions to ensure they understand the story. The change-of-location task is often used to examine children's FB

(Wellman et al., 2001). Research has found that most children who are four years of age pass these tasks; however, most younger children fail these tasks (Baron-Cohen et al., 1985; Clements & Perner, 1994; Wellman et al., 2001). Therefore, children, from about four years of age, show their understanding that others can have FBs.

Why do Younger Children Fail Change-of-Location Tasks?

It has been proposed that people have a *reality bias*, where they predict how others will behave based on the reality of where the objects are actually located at the time. This may occur because individuals do not have a ToM or they do but are not using it (Keysar, Lin, & Bar, 2003). Some research has not found support for the reality bias in young children (i.e., two- and three-year-olds) or adults (Southgate, Senju, & Csibra, 2007; Wang & Leslie, 2016). Another account, the *true-belief default*, suggests that people attribute their own beliefs to others because their own beliefs are true most of the time (Leslie, Friedman, & German, 2004). Wang and Leslie (2016) found support for the true-belief default in young children and adults, but Rubio-Fernandez (2017) failed to find support in adults. Other accounts have been offered to explain why younger children fail change-of-location tasks.

Conceptual change and *early competence* are general accounts argued to answer why younger children fail change-of-location tasks. According to conceptual change, children's concept of others shifts to enable these children to understand others' mental states (Wellman et al., 2001). Children, who lacked this ability, can now understand when other people's mental representations of the world represent or misrepresent the actual state of the world (Perner, 1991). In support of conceptual change, Wellman et al.'s (2001) meta-analysis indicated that children's performance in change-of-location tasks and other similar tasks changes from two and a half years to five years of age. According to early competence, young children can conceptualize other individuals' mental states. These tasks do not reveal children's ToM ability because these children lack the cognitive skills (e.g., attention) to meet the other required task demands. These children also have a limited ability to process information (Wellman et al., 2001). However, Wellman et al. (2001) found that young children gave correct responses in change-of-location tasks 50% or less of the time, even when these tasks had few cognitive demands.

While accounts vary in terms of early competence or conceptual change, they also vary if they are *domain-general*, or *domain-specific* (Cohen & German, 2010). In particular, accounts attribute individuals' ability to interpret and infer others' behaviour, to a greater extent, to either general or specialized "cognitive mechanisms," respectively. It has been proposed that domain-general cognitive processes enable people to infer different types of content including those related to mental states. For instance, domain-general cognitive processes may enable individuals to recognize when content that other individuals can observe (e.g., maps, signs, photos, and arrows) is inconsistent with reality. Conversely, it has been theorized that domain-specific-processes enable individuals to interpret and infer others' mental states based on their behaviour with an exclusive "cognitive mechanism." Leslie and colleagues (Leslie, Friedman, & German, 2004; Leslie & Polizzi, 1998), for instance, proposed that people have a *ToM mechanism* (ToMM) in their brain that enables them to attribute others' beliefs and desires. Cohen and German's (2010) findings support that specific mechanisms are involved in ToM because participants responded more quickly to mental states (e.g., beliefs) versus general representations (e.g., maps or arrows).

In summary, many theories have been proposed to explain why younger children fail change-of-location tasks, including the reality bias and the true-belief default account. These theories vary in research support (e.g., Rubio-Fernandez, 2017). It is unclear whether the reality bias is considered an early competence or conceptual change account, but it involves a domain-specific process. The true-belief default account is considered an early competence, domain-specific process by some researchers (e.g., Cohen & German, 2010; Leslie et al., 2004). Ultimately, Wellman et al. (2001) found support for conceptual change but not early competence accounts.

Do People Have ToM by Age Four or Earlier?

ToM has been investigated in children younger than four years of age. This research used change-in-location tasks that were modified to study ToM indirectly so that the verbal demands were absent or decreased (Wang & Leslie, 2016). Some studies used a video camera or an eye tracker to measure *anticipatory looking*, or where children looked when cued to anticipate characters' behaviour in change-of-location tasks (Garnham & Ruffman, 2001). Research found that young children looked where the characters, with FBs, believed the object was located. Thus, these 25-month-olds (Southgate et al., 2007) as well as 35- to 53-month-olds (Clements & Perner, 1994) correctly anticipated the location that the characters would look for the object. These studies suggest that young children understand FBs. Therefore, some researchers question whether children have ToM before age four (e.g., Clements & Perner, 1994; Kovács, Téglás & Endress, 2010; Onishi & Baillargeon, 2005; Southgate et al., 2007; See Perner & Ruffman, 2005 and Ruffman, 2014 for an alternative interpretation).

These discrepant findings have been explained by distinguishing between *implicit* (not verbally expressed) and *explicit* (verbally expressed) understanding of others' beliefs. The change-of-location task examines explicit ToM if participants need to respond to questions regarding characters' beliefs or behaviours (Clements & Perner, 1994). However, tasks examine implicit ToM if: (1) the beliefs of others are not referred to and (2) predicting characters' actions by considering their beliefs is not necessary (Kovács et al., 2010). Therefore, implicit ToM can be shown through anticipatory looking. Thus, the results from Clements and Perner (1994) suggest that implicit ToM is a distinct "type of knowledge," and it develops before explicit ToM, which may still develop by four years of age. Further, Onishi and Baillargeon (2005) provided evidence that infants can understand that others' beliefs may accurately or inaccurately represent reality, on a basic and implicit level at minimum.

To summarize, research has examined ToM in children with direct and indirect measures. This has resulted in discrepant findings and subsequent theories to account for these findings. Children four years of age and older pass change-of-location tasks that study ToM with direct measures (Wellman et al., 2001). Meanwhile, younger children pass these tasks with indirect measures of ToM (e.g., Kovács et al., 2010). Further, it has been proposed that people have an implicit and explicit ToM, which has been studied with indirect and direct measures, respectively (Clements & Perner, 1994). Thus, people may have ToM earlier than four years of age.

What are the Characteristics of Implicit and Explicit ToM Processes?

It has been proposed that people have implicit and explicit ToM processes that enable them to unintentionally and intentionally follow others' mental states, respectively (Schneider, Slaughter, & Dux, 2015). Various theories state that people have one ToM process that is fast and has reduced cognitive demands. People have another process that is slow and has greater cognitive demands (Schneider et al., 2015). Of particular interest is Apperly and Butterfill's (2009) *two-systems account* that individuals have one resource-efficient, yet "limited and inflexible" ToM process. There is also a "flexible" ToM process, which requires language, executive function and other general cognitive resources. Broadly speaking, *cognitive flexibility* enables people to take a different point of view and make quick modifications when a situation has changed (Diamond, 2013). Additionally, *executive functions* enable individuals to think through ideas and future actions, embrace challenges that are new and unexpected, respond to temptations with self-control, as well as pay attention to certain things while ignoring others (Diamond, 2013). Young children use the first process to pass several ToM tasks. Thereafter, young children undergo development in their general cognitive abilities, so they develop the second process when they are older. Adults can use both processes simultaneously. The second process allows adults to use top-down processing when interacting and communicating with others (e.g., when a speaker considers their audience's prior knowledge before a speech) as well as to explicitly consider the source of mental states and how to justify them.

Previous research supports that people have implicit ToM processes that can be observed over a prolonged period and across multiple trials (Schneider, Bayliss, Becker, & Dux, 2012; Schneider, Lam, Bayliss, & Dux, 2012). In an implicit change-of-location task, participants looked longer, in anticipation, at the location without the object in the FB versus TB (i.e., when the character and participant saw the object move) trials (Schneider, Bayliss, Becker, & Dux, 2012; Schneider, Lam, et al., 2012). In the same studies, most participants accurately performed the filler tasks and did not report explicitly tracking the actor's belief, suggesting that people have implicit ToM processes that function without their awareness.

Bargh (1994) proposed characteristics for automatic mental processes which support the automaticity of implicit ToM processes. Bargh proposed that mental processes are automatic, depending on how conscious individuals are of their mental processes, how intentional it is for individuals to control when their mental processes begin, how controllable it is for individuals to end their mental processes that have already begun, and how efficient it is on individuals' cognitive resources to perform the mental processes. Research found that implicit ToM processes are unconscious, as well as unintentional and uncontrollable (e.g., Schneider, Bayliss, Becker, & Dux, 2012; Schneider, Nott, & Dux, 2014). However, some research suggested that implicit ToM processes are not efficient with cognitive resources (Wang & Leslie, 2016), which contradicted Apperly and Butterfill's (2009) proposed resource-efficient process. For example, Schneider et al. (2014) found that adult participants, completing an implicit change-of-location task where they were told to follow an object's location, did not look longer at the location without the object in the FB versus TB condition, indicating cognitive resource inefficiency. This suggests that implicit ToM processes may be largely affected by top-down processing.

Ultimately, various characteristics of implicit and explicit ToM processes have been proposed, for example, in the two-systems account (Apperly & Butterfill, 2009). Most of these characteristics have been supported with research findings. It has been shown that people use their implicit ToM processes over a prolonged period and across multiple-trials, and that these processes are automatic (e.g., Schneider, Bayliss, Becker, & Dux, 2012). Overall, research offers and supports many proposed characteristics of both ToM processes.

Is it Automatic for People to use Their ToM When Observing Others' Behaviour?

Research has studied whether people infer others' beliefs before or only after others behave in a certain way and people have to consider others' beliefs (Schneider et al., 2015). Apperly et al.'s (2006) findings indicated that adults do not automatically attribute others' beliefs. Apperly et al. found that participants processed the character's belief only when it was required in the change-of-location task. Other researchers' findings, however, suggest that individuals automatically encode and represent others' mental states (e.g., Kovács et al., 2010). Kovács et al. (2010) found support that adult participants automatically track the beliefs of a character in animated movies, when participants were only instructed to indicate the presence of an object. Similarly, Cohen and German (2009) found that participants processed a character's belief, even though following where an object was located was the only requirement of the task. Nevertheless, studying automatic processes in ToM is complex. While some researchers study ToM like it is one process that is automatic or non-automatic, other researchers (German & Cohen, 2012) study what is required for people to respond with automatic ToM processes.

Are Implicit and Explicit ToM Processes Distinct From Each Other?

According to Schneider et al. (2015), there is considerable research to support that implicit and explicit ToM processes are distinct from each other. This evidence includes studies indicating that adults with Autistic Spectrum Disorder (ASD) have an implicit ToM deficit yet demonstrate explicit ToM (Schneider, Slaughter, Bayliss, & Dux, 2013; Schuwerk, Vuori, & Sodian, 2015; Senju, Southgate, White, & Frith, 2009). That is, studies found that neurotypical and ASD adults passed most of the explicit ToM tasks. Only neurotypical adults, however, looked longer and/or first looked, in anticipation, at the location without the object when an actor had a FB. Additionally, there is evidence that children's and adults' implicit ToM processes have *blind spots* or are inaccurate when automatically and quickly attributing FB about an object's identity (i.e., whether an object is red or blue; Low & Watts, 2013). Schneider et al. (2015), however, suggest that there is "overlap" in implicit and explicit ToM processes, considering that they both seem to require working-memory resources. There is evidence, with studies using the implicit change-of-location task, that implicit ToM processes require executive functions to some extent (Schneider, Lam, et al., 2012; Schneider et al., 2015; Wang & Leslie, 2016). Therefore, these results contradict Apperly and Butterfill's (2009) proposal that there are resource-efficient processes that do not require executive functions. Thus, implicit and explicit ToM could be distinct processes that involve similar executive functions.

Can Adults' Private Knowledge Affect One's Ability to Understand Others' Beliefs?

Studies have shown that adults' ability to understand others' beliefs can be biased by their own private knowledge; thus, adults can make ToM errors (Sommerville, Bernstein, & Meltzoff, 2013). For instance, adult participants made more ToM errors when they thought another person was asking them to move an object only they could see instead of an object they and the other person could see (Keysar, Barr, Balin, & Brauner, 2000). In particular, participants' behavioural response and eye movements revealed a delay in identifying and moving the correct object. There were similar results when participants knew the contents of a bag, but the other person did not know or had a FB about the bag's contents (Keysar et al., 2003). Dumontheil, Apperly, and Blakemore (2010) also found that many participants did not consider the perspective of another person; however, there were more accurate responses in adults versus late adolescent participants. Further, these findings suggest that it took less time for participants who considered a person's perspective to inhibit moving the incorrect object in favor of moving the correct object, compared to participants who followed a rule regarding which objects to ignore.

Several theories have been proposed to explain the reason for adults' ToM errors. These theories vary in how involved adults' executive functions are in their ToM errors. Some researchers (Keysar et al., 2000; Keysar et al., 2003) suggested that participants made ToM errors because they reasoned that other people would identify and correct any of their misunderstandings; thus, these ToM errors may use less of their cognitive resources. Further, Rubio-Fernandez and Glucksberg (2012) found that monolingual adults, compared to bilingual adults made more ToM errors. In an explicit change-of-location task, significantly more bilingual versus monolingual adults first looked correctly at the location without the object in the FB trial. In contrast, most adults, monolingual and bilingual, first looked correctly at the location with the object in the TB trial. Rubio-Fernandez and Glucksberg (2012) referred to improved executive functions (i.e., the ability to pay attention to one language without having their other language interfere) as one explanation for why bilingual versus monolingual adults may have an "advantage" in FB processing.

Dumontheil et al. (2010) proposed a developmental trajectory of their participants, where ToM and executive functions interacted, and developed together from childhood to adulthood. From childhood to mid-adolescence, participants developed the necessary *working memory* and *inhibitory control* required in the task. Working memory is the brain system that enables individuals to temporarily store and manipulate information that is required in complex cognitive tasks like understanding language, learning, and logical thinking (Baddeley, 1992). Further, inhibitory control enables people to stop an intended action or an action that has started (van Gaal, Ridderinkhof, Fahrenfort, Scholte, & Lamme, 2008). According to Dumontheil et al. (2010), late adolescent participants had the working memory and inhibitory control, like adults, that was required to take others' perspectives. Still, late adolescent participants had to further develop their ability to respond appropriately after taking others' perspectives. Particularly, Dumontheil et al. (2010) argued that from late adolescence to adulthood, participants developed the tendency to use their perspective taking ability more often to understand how others behave. This proposal related to Keysar et al.'s (2003) argument that adults use their ToM similarly to how they use a "tool." Adults can use their ToM when they consciously reason about others' actions, but it has not become their common and unconscious way of interpreting others' actions.

Overall, research found that one's private knowledge affected one's ability to understand others' beliefs. Studies showed that children, adolescents, and adults as well as monolingual adults made ToM errors (Dumontheil et al., 2010; Rubio-Fernandez & Glucksberg, 2012). These researchers proposed several reasons for an adult to make ToM errors. Overall, people's private knowledge can hinder ToM in childhood and adulthood.

Is There a Continuous Measure of ToM and ToM Errors?

Change-of-location tasks typically have two distinct locations where the object can be put. Therefore, when individuals are asked where a character will look for an object, they either fail or pass depending on if they respond with the correct or incorrect location in line with a character's beliefs. Thus, this task measures whether people can or cannot understand that it is possible for others to have FBs (Baron-Cohen et al., 1985). The novel *Sandbox* task is a *continuous* ToM measure (Sommerville et al., 2013). It measures the extent to which people's private knowledge affects their ability to understand others' beliefs (i.e., make ToM errors). In the Sandbox task, participants hear several stories, based on the Sally-Anne narratives. The experimenter puts real objects in locations, within a 5-foot-long (152.4 cm) box containing Styrofoam peanuts, as he or she reads the stories aloud. In the stories, a main character puts an object in one location and exits the scene. In his or her absence, a secondary character puts the same object in a second location. After completing a filler task, participants have to point to where the main character, who returns, would look for the object. In prior studies, participants have been either in the FB condition (they indicated where the main character, with a FB, would look for the object), the *memory control* condition (they indicated where the object was located originally; Sommerville et al., 2013), or in both conditions (Coburn, Bernstein, & Begeer, 2015; Mahy, Bernstein, Gerrard, & Atance, 2017).

Participants' responses are measured with a ruler, in centimeters or millimeters. Their ToM error is calculated by comparing participants' responses to the object's actual location (Mahy et al., 2017). If they respond closer to the object's actual location in the FB versus memory control condition, participants show that they have used their own private knowledge (i.e., the object's actual location) to predict the main character's behaviour instead of considering the character's naïve perspective (Coburn et al., 2015). Thus, participants have to set aside their own knowledge and possibly the "pull of reality" (Bernstein, Thornton, & Sommerville, 2011) because the object's actual location is more salient and current than the object's first location (Bernstein, Coolin, Fischer, Thornton, & Sommerville, 2017). Further, the memory control condition checks whether participants recall the object's first location (Coburn et al., 2015). Another version of the Sandbox task is performed with paper and pencil, and it has convergent and discriminant validity (Begeer, Bernstein, van Wijhe, Scheeren, & Koot, 2012; Coburn et al., 2015; Mahy et al., 2017).

The Sandbox task has been used to study how people's ability to consider others' perspective develops through their lifespan (Bernstein et al., 2017). Sommerville et al. (2013) found that young children's (three- and five-year-olds) and adults' knowledge of where the object was actually located biased their responses in that direction, in the FB condition. Participants made fewer ToM errors in the memory control condition. Further, the number of ToM errors decreased as the age of participants increased such that children made more ToM errors compared to adults. Sommerville et al. found that five-year-olds outperformed three-year-olds on an explicit change-of-location task, and those that passed this task made fewer ToM errors in the Sandbox task. Further, Mahy et al. (2017) found a negative correlation between age and ToM errors in the Sandbox task in three- to seven-year-olds, showing that participants' comprehension of FB increased with age. These findings support that the Sandbox task can detect children's growing ability to understand FB (i.e., shown by making fewer ToM errors) as they become older. Further, children's responses to the FB condition in the Sandbox and an explicit change-of-location task correlated significantly, independent of memory, age, and vocabulary. This suggests that the tasks resemble each other regarding the concept the tasks measure as well as the tasks' "structure". In addition, Bernstein et al. (2017) found that young children (i.e., three- and five-year-olds), school-age children, and young adults made comparable ToM errors.

Sandbox task studies found that middle-aged and older versus younger adults made more ToM errors (e.g., Bernstein et al., 2011). Bernstein et al. (2017) found that older compared to younger adults made more ToM errors in the FB versus memory control condition. This was a significant, although "small" effect, indicating that the ability to understand others' FBs differed only modestly between these age groups. Likewise, in Bernstein et al.'s (2011) study, middle-aged and older versus younger adults made more ToM errors in the FB condition. This was found independent of participants' "language ability, executive functioning, processing speed or memory" (Bernstein et al., 2011, p. 500). Similarly, one meta-analysis found that older compared to younger adults had greater ToM difficulties in various ToM tasks (i.e., change-of-location task) with different domains (i.e., cognitive) and modalities (i.e. verbal, visual). These age-related difficulties were higher in FB versus control tasks. This indicates that older adults have a ToM deficit that cannot simply be accounted for by "non-ToM-related task demands" (Henry, Phillips, Ruffman, & Bailey, 2013, p. 832).

The Sandbox task has several advantages over the change-of-location task. While testing older children and adults with the change-of-location task results in ceiling effects, it is possible to test people of many ages with the Sandbox task (Coburn et al., 2015). Further, Sommerville et al. (2013) note that individuals frequently do not remember accurately where an object was located initially or had been moved. Therefore, the Sandbox task is ecologically valid as a continuous measure, compared to the change-of-location task in which participants choose one of two discrete locations (Sommerville et al., 2013). Overall, the Sandbox addresses some shortcomings of the change-of-location task.

To summarize, the Sandbox task is a continuous measure of ToM (Sommerville et al., 2013). Using the Sandbox task, studies have found that people of various ages make ToM errors (e.g., Bernstein et al., 2017). Further, the Sandbox task is more ecologically valid than the change-of-location task (Sommerville et al., 2013). Overall, the Sandbox task can measure ToM and ToM errors continuously throughout people's lifespan (e.g., Bernstein et al., 2017).

The Current Research

The current research aimed to show that adults can understand that others' beliefs can be false (Wimmer & Perner, 1983). Still, adults can make ToM errors when their private knowledge hinders their ability to understand others' beliefs (Sommerville et al., 2013). The current work used an eye tracker to replicate Cho and Cohen's (2017) study that used a novel implicit change-of-location task (herein referred to as the *Implicit ToM task*). In this task, participants watched several videos and performed a filler task. Cho and Cohen found that when adult participants watched videos depicting change-of-location scenarios, they looked longer, in anticipation, at the location that a character would search for an object. That is, when a character returned to a room, participants looked at a box without and with an animal, in the FB and TB conditions, respectively. Participants had no reason to follow a character's beliefs because they were performing a filler task. Therefore, the Implicit ToM task studied their implicit ToM processes. Additionally, Cho and Cohen's study included eye tracking. The current study also attempted to extend Cho and Cohen's study by examining whether participants first look at the location that a character would search for an object.

Further, in the current study, participants completed the Sandbox task, which is an explicit change-of-location task (herein referred to as the *Explicit ToM task)*. In the computerized Explicit ToM task (adapted from Sommerville et al., 2013; Derksen, 2017), participants heard several stories. In these stories, participants answered a *search* question and a *memory* question to indicate, respectively, where a character would look for an object or where a character put an object. The eye-tracker recorded where participants looked longer in the Explicit ToM task. This was the first study to use eye tracking in the Explicit ToM task.

Method

Participants and Sampling Procedures

The current study recruited 70 adult participants from Kwantlen Polytechnic University (KPU) (80.9%) and the public (19.2%). Participants' age ranged from 18 to 44-years-old (M = 22.39, SD = 4.60, 72.9% women and 27.1% men). Participants described themselves as White (28.6%), Indian (27.1%), Asian (32.9%), Black (2.9%), Hispanic (1.4%), and Other (7.1%). Nearly half the participants' (48.6%) first language was English, followed by Punjabi (31.4%), Hindi (5.7%) and others (14.3%). These participants, however, self-reported that they spoke English very (66.7%) or moderately well (33.3%), wrote English very (80.6%) or moderately well (19.5%), and used English every day (97.2%) or every few days (2.8%). The study excluded people whose eye movements could not be detected and/or who could have had a seizure triggered by the eye tracker. KPU students and employees were recruited through the KPU Research Pool and posters posted at KPU campuses. Participants from the community were recruited through ads posted on Facebook, Kijiji, and Craigslist. Participants received course credit or 10 Canadian dollars.

A power analysis using G-Power*3 computer program (Faul, Erdfelder, & Lang, 2007) indicated that 130 participants were needed to obtain a medium effect size (w = 0.3), with a Goodness-of-fit tests: Contingency tables test. Further, 134 participants were necessary to obtain an intermediate effect size (d = 0.30), with a Means: Difference between two dependent means (matched pairs) two-tailed test. Finally, 88 participants were needed to obtain an intermediate effect size (f = 0.15), for an ANOVA: Repeated measures, within-between interaction test with two groups and four measurements. As such, it was necessary to recruit 134 participants. The average sample size of published studies on adult eye tracking examining ToM is 57 adult participants (Kovács et al., 2010; Low & Watts, 2013; Rubio-Fernandez & Glucksberg, 2012; Schneider, Bayliss, Becker, & Dux, 2012; Schneider, Lam, et al., 2012; Schneider et al., 2013; Senju et al., 2009; Wang & Leslie, 2016).

This study was approved by the KPU research ethics board. Further, detailed preregistration, including hypotheses and analyses, were archived at open science framework (Giesbrecht, Bernstein, & Derksen, 2018, February 19, osf.io/ft3ax). This report is based on a partial n because data has only been collected for 73 out of 134 participants.

Materials

Implicit ToM task. The current study used the Implicit ToM task from Cho and Cohen (2017) to replicate and extend their findings. The task was displayed on the full-screen of computer monitors (1920px x 1080px). An example of the estimated size of the scenarios is shown in Figure 1.

The current study used the same procedures, including the same filler task, to replicate Cho and Cohen's (2017) study[2]. Participants were asked to watch animated characters put an animal in one of two boxes, exit the room, and come back to retrieve the animal (I. Cho, personal communication, November 3, 2017). Then, participants were told to indicate the location of the character, who stood behind the right or left-hand box, by pressing the right or left arrow key, respectively (see Figure 2; I. Cho, personal communication, November 3, 2017). The latter was a filler task that did not require participants to follow the character's belief about where the animal was located. As a manipulation check, there was a comparison between participants' correct responses (i.e., number of times they pressed the correct computer key) in the FB and TB conditions; this will be returned to later.

The Implicit ToM task consisted of 40 videos showing different characters putting different animals in one of two boxes. Half the videos were in the FB condition (see Figure 2), where the character exited the room. In the character's absence, the animal moved into the second box. The remaining videos were in the TB condition (see Figure 2), where the character exited and re-entered the room, and then saw the animal move into the second box. Then, in all the videos, the character returned to the room to retrieve the animal (I. Cho, personal communication, November 3, 2017).

If the character has a FB or TB, participants should look longer and first look, in anticipation, to the first or second box, respectively. That is, these were *correct locations* because the character would look for the animal in these boxes, based on their FB or TB. Conversely, the *incorrect location* was where the character would not look for the animal because they believed that the animal was located in the other box. It was randomized whether the correct or incorrect location was the box to the left or right of the character. There were two *areas of interest*

[2] Replicating Cho and Cohen's study would confirm three things: (1) the eye tracker was working as it should, (2) participants were responding as they should, and (3) the experimenter was conducting this study as s/he should. If Cho and Cohen's study was replicated and the Explicit ToM task yielded null eye-tracking effects, then it could be concluded that the Sandbox's null effects were not due to the three aforementioned things. Thus, these three confirmations would inform the interpretations that could be made about the Explicit ToM task's findings.

(specified sections in the task for the eye tracker to record eye movement during a given time), one around the correct location, and another around the incorrect location. These two areas of interest included 30 pixels extending beyond these locations to follow Cho and Cohen's study (I. Cho, personal communication, February 2, 2018). The eye tracker measured whether participants looked longer and/or first looked at the correct or incorrect location, in anticipation of the character's behaviour.

It is worth noting that participants could have looked at two discrete locations in the Implicit ToM task because this task was a modified change-of-location task with two discrete locations. Conversely, participants in the Explicit ToM task could have looked towards a range of locations within a box because this task was a continuous space (Sommerville et al., 2013). Additionally, because the Implicit ToM task measured implicit ToM, participants had to complete it first, before they completed the Explicit ToM task, which was an explicit ToM task. If participants performed the Explicit ToM task first, it is possible that they would have been more likely to explicitly follow a character's beliefs in the Implicit ToM task.

Explicit ToM task. The current study used a computerized Explicit ToM task (adapted from Derksen, 2017) based on Sommerville et al.'s (2013) original Explicit ToM task. The task was displayed on the full-screen of computer monitors (1920px x 1080px). The size of the box was 1600px to 170px, and the distance between the initial location that the main character put an object and a second location that a secondary character moved the object was 450px. An example of these dimensions is shown in Figure 3. The object's first location was semi-random as it may have been in any location within the box except in locations where the object's first or second location was near the box's edge. These boundaries are depicted with a dotted line as shown in Figure 4. In the task, participants heard stories based on the change-of-location task, where a character either had a FB or TB about an object's location (see Figure, 4, 5, and 6). Then, participants were asked a search question, "Where will the main character look for the object?" (see Figure 4 and 5) and a memory question, "Where did the main character put the object?" (see Figure 6). The order in which the two questions were asked was counterbalanced to prevent people from developing a strategy when answering the questions. Further, participants answered these questions by using a mouse to click inside a picture of a box, which represented an actual box. There were 12 trials in total: Six trials were in the FB condition, and six trials were in the TB condition (see Appendix A). FB and TB trials were randomized.

This research was the first to use an eye tracker with the Explicit ToM task. The primary measure of interest was where participants looked when the main character returned. This included the locations on the screen in which they looked longer, in anticipation of the main character's behaviour, starting when the search or memory question began with the first word, "where," until participants responded to the question. While prior research using implicit change of location tasks had a set time to measure participants' anticipatory behaviours (Schneider, Bayliss, Becker, & Dux, 2012), the current study was interested in examining where participants looked from the time they heard the question until they responded to it. Participants were expected to look longer towards the location where the secondary character moved the object when answering the search question in the FB condition. In contrast, participants were expected to look shorter towards this location when answering the search question in the TB condition and the memory question in both conditions.

Participants' ToM errors were measured. ToM error scores for the FB and TB condition were calculated in pixels. As participants completed the Explicit ToM task, the computer recorded their explicit responses to the search and memory question. The Explicit ToM task program recorded participants' responses, the location that they indicated in response to the question. The program also recorded the correct location, the location that was the correct response to the question. Specifically, ToM error scores were computed by calculating the distance between the response location and the correct location. Participants' ToM error scores were a positive value if their response location was towards, or in the same direction as, the incorrect location (see Figure 3). Conversely, they were a negative value if their response location was away from, or in the opposite direction as, the incorrect location (Sommerville et al., 2013).

Funneled Debriefing Procedure. The current study used a computerized funneled debriefing procedure (see Appendix B). Schneider, Bayliss, Becker, and Dux (2012) adapted this procedure from Bargh and Chartrand (2000) to examine whether participants explicitly followed an actor's beliefs in an implicit change-of-location task. It was sensitive enough to detect participants that may have been aware that the study was investigating implicit cognitive processes. These participants were excluded in Schneider, Bayliss, Becker, and Dux (2012) study's analysis. Since then, other studies examining implicit ToM processes have used the funneled debriefing procedure (Schneider, Lam, et al., 2012; Schneider et al., 2013; Schneider et al., 2014).

The Funneled Debriefing Procedure included seven open-ended questions, such as "What do you think the purpose of the experiment was?" (Schneider, Bayliss, Becker & Dux, 2012, p. 438). Some of the questions referred to a filler task, called the "tone detection task," that Schneider, Bayliss, Becker, and Dux (2012) included in their study. The funneled debriefing procedure was modified for the current study by referring to the filler task, in the Implicit ToM task, called the "task to detect where a character is standing." If their answers suggested that they may have been explicitly processing the character's belief, it would influence the current study's ability to draw conclusions about participants' implicit ToM processes. All participants were included in the statistical analyses for four hypotheses. Then, as exploratory analyses, the same analyses were run with participants who reported that they were unaware of the character's beliefs, to see whether there were any differences in the results. Participants completed the funneled debriefing procedure after the Implicit ToM task because it had specific questions related to that task.

Procedure

The experiment took 50 to 60 minutes to complete. Participants were verbally informed and signed a consent form. Then, they were instructed to remove any eye makeup (makeup remover and cotton pads were available upon request). Participants were able to adjust the height of their chair as well as place their head in a chin rest in a comfortable position. The chin rest was located 59 cm from the computer monitor. When ready, participants followed the instructions of the experimenter to calibrate a Tobii Pro X2-60 eye tracker. If calibration was poor, participants wearing glasses were asked to clean their glasses, and then to remove their glasses (if they could see the computer well without glasses). Participants wearing contacts were also asked if they would remove their contacts (if they could see the computer well without contacts). If participants wearing glasses or contacts were not willing to remove them, the experimenter attempted to recalibrate and have participants begin the tasks. If the second calibration failed, participants skipped the Implicit ToM task and funneled debriefing procedure and completed the Sandbox.

Participants completed the Implicit ToM task, followed by the funneled debriefing procedure. Then, participants were offered an optional 5-minute break. If they chose to take a break, they were instructed, upon their return, to find a comfortable position and calibrate the eye tracker again. Next, participants completed the Sandbox. Afterwards, all participants completed a demographic questionnaire, which asked general information, for example, about their age. Finally, the experimenter verbally debriefed participants.

Experimental Design. The current study has two ToM tasks: an Implicit and Explicit ToM task. The Implicit ToM task followed a one-factor (belief type: false, true) within-subjects experimental design. The Explicit ToM task followed a 2 (belief type: false, true) x 2 (question type: search, memory) within-subjects experimental design. In the Implicit ToM task, the independent variable was belief type and the dependent variables were *dwell time* (duration of look) and first look. In the Explicit ToM task, the independent variables were belief and question type, and the dependent variables were dwell time and participants' ToM error scores. Participants had to look at one of the areas of interest for at least 100 milliseconds for that eye movement to be included in the dwell time and first look analyses (based on Rubio-Fernandez & Glucksberg, 2012). Because this study involved multiple comparisons, a conservative alpha level, $p < .01$, was set. This alpha level was chosen to balance Type I and Type II error. Given the multiple comparisons, this study's results also require replication.

Hypotheses

Hypothesis 1 and 2 (Implicit ToM task). Based on prior work (Low & Watts, 2013; Schneider, Bayliss, Becker, & Dux, 2012; Schneider, Lam, et al., 2012; Schneider et al., 2013; Schneider et al., 2014; Schuwerk et al., 2015; Senju et al., 2009), in Hypothesis 1 it was hypothesized that there would be a greater proportion of dwell time to the correct versus the incorrect location in the FB condition. Additionally, Hypothesis 2 stated that there would be a greater proportion of dwell time to the correct versus the incorrect location in the TB condition. Therefore, this study aimed to replicate Cho and Cohen's (2017) findings that adult participants looked longer, in anticipation, at the correct location that a character with a FB would look for an animal. Previous research, which involved implicit change-of-location tasks, found that neurotypical adult participants looked longer in anticipation at the correct location when a character returned to a scenario in the FB and TB condition (Low & Watts, 2013; Schneider, Bayliss, Becker & Dux, 2012; Schneider, Lam, et al., 2012; Schneider et al., 2013; Schneider et al., 2014; Schuwerk et al., 2015; Senju et al., 2009).

Hypothesis 3 and 4 (Implicit ToM task). Based on prior work (Low & Watts, 2013; Schuwerk et al., 2015; Senju et al., 2009), in Hypothesis 3 it was hypothesized that there would be a greater proportion of first looks to the correct versus the incorrect location when a character had a FB about an animal's location (FB condition). Furthermore, Hypothesis 4 stated that there would be a greater proportion of first looks to the correct versus the incorrect location when a character had a TB about an animal's location (TB condition).

The current study extended Cho and Cohen's findings by testing whether participants first looked, in anticipation, to the correct location. Prior research with implicit change-of-location tasks found that when a character returned, neurotypical adult participants first looked, in anticipation, at the correct location in the FB and TB condition (Low & Watts, 2013; Schuwerk et al., 2015; Senju et al., 2009).

Hypothesis 5 (Explicit ToM task). Based on prior work (Begeer et al., 2012; Sommerville et al., 2013), it was hypothesized that participants would make more ToM errors when they answered (1) the search question in the FB condition. In comparison, participants would make fewer ToM errors when answering (2) the search and (3) memory question in the TB condition, as well as (3) the memory question in the FB condition. Firstly, it was only expected that participants would make more ToM errors when they were asked the search question versus the memory question in the FB condition. When participants answered the search question, they were required to reason about the main character's FB. Participants knew the object was no longer located where the character thought it was (i.e., where they initially put it). Therefore, in the search question, participants had to ignore their own private knowledge of the object's actual location and consider the character's naïve perspective or FB (Coburn et al., 2015) to respond towards the correct location. If they did not consider the character's FB, which was expected, participants' responses would be towards the incorrect location. Secondly, when participants were asked the search question in the TB condition, participants and the character knew where the object was really located. Therefore, in the search question, participants' responses should be towards the correct location whether they considered the character's perspective or not. Thirdly, asking participants the memory question, in the FB or TB condition, checked participants' memory for the object's initial location (Coburn et al., 2015). Thus, in the memory question, participants' responses should be towards the correct location. If participants passed the memory check in the FB and TB condition, it was possible to conclude whether participants showed difficulty reasoning about a character's FB.

Prior research with the Explicit ToM task showed that adult participants made more ToM errors in the FB compared to the memory control trials. There was also one TB trial in prior research with the Explicit ToM task, but it was not analyzed. Adult participants made fewer ToM errors than children; however, this task was sensitive enough to detect a bias in adults and adolescents (Begeer et al., 2012; Sommerville et al., 2013).

Results

Participants' anticipatory looking was measured for 5 seconds, starting when the character returned to the room to when s/he was about to move towards one of two boxes (I. Cho, personal communication, November 3, 2017). Participants were included (*N* = 57) if they looked to the correct location, incorrect location, or the character, at least once in 80 percent of the trials; this was based on Rubio-Fernandez and Glucksberg's (2012) selection criterion. Thirteen participants were excluded because they did not meet this criterion.

In the Explicit ToM task, it was a simple process to calculate participants' ToM error scores. It was an extensive process, however, to collect their dwell time data from the eye tracker. In the ToM error score and dwell time analyses, two participants were excluded because they had missing output excel files. Additionally, in the dwell time analysis, participants were included (*N* = 45) if they looked towards the correct and/or incorrect location, at least once in 80 percent of the trials; this was based on Rubio-Fernandez and Glucksberg's (2012) selection criterion. Twenty-three participants were excluded because they did not meet this criterion.

For the transparency of the following statistical analyses, it is worth noting that 10 participants took off their glasses because they indicated that they could see the computer without them. Three and one of these participants did not meet the selection criterion for the Implicit ToM or Explicit ToM task, respectively. Additionally, 15 participants wore their glasses because they indicated that they could not see the computer without them. Two and five of these participants did not meet the selection criterion for the Implicit ToM or Explicit ToM task, respectively.

Hypothesis 1 (Implicit ToM task): μ (FB Proportion of Dwell Time Correct) > μ (FB Proportion of Dwell Time Incorrect)

A paired samples *t*-test was conducted to evaluate the proportion of dwell time to the correct and incorrect location in the FB condition; higher proportions indicate greater dwell time. Proportion of dwell time to the correct and incorrect location was calculated as follows: Dwell time to the correct location divided by dwell time to the correct and incorrect location equaled proportion of dwell time to the correct location; dwell time to the incorrect location divided by dwell time to the correct and incorrect location equaled proportion of dwell time to the incorrect location. Prior to conducting the analysis, the assumption of normally distributed difference scores were examined. The assumption was considered satisfied, as the skew and kurtosis levels were estimated at 2.08 and 4.63, respectively, which was near or less than the maximum allowable values for a *t*-test (skew < |2.0| and kurtosis < |9.0|; Posten, 1984). Further, the correlation between the two conditions was estimated at $r = -1.00$, $p < .001$, suggesting that the dependent samples *t*-test is appropriate in this case. Additionally, paired samples *t*-tests revealed non-significant differences when the correct, $t(56) = -1.23$, $p = .22$, and incorrect location, $t(56) = 1.03$, $p = .31$, were on the left or right-side. Thus, the data was collapsed together when the correct and incorrect location was the left or right-side box. Contrary to the hypothesis, there was a significantly greater proportion of dwell time to the incorrect ($M = .84$, $SD = .19$) compared to the correct ($M = .16$, $SD = .19$) location, when a character had a FB, $t(56) = -13.13$, $p < .001$, Cohen's $d = -1.74$. A graphical representation of the means and adjusted 99% confidence intervals is shown in Figure 7.

Hypothesis 2 (Implicit ToM task): μ (TB Proportion of Dwell Time Correct) > μ (TB Proportion of Dwell Time Incorrect)

Paired samples *t*-tests were conducted to evaluate the proportion of dwell time to the correct and incorrect location in the TB condition; higher proportions indicate greater dwell time. Proportion of dwell time to the correct and incorrect location was calculated the same way as the previous analysis. Prior to conducting the analysis, the assumption of normally distributed difference scores was examined. The assumption was considered satisfied, as the skew and kurtosis levels were estimated at -1.43 and 1.76 respectively, which was less than the maximum allowable values for a *t*-test (skew < |2.0| and kurtosis < |9.0|; Posten, 1984). Further, the correlation between the two conditions was estimated at $r = -.83$, $p < .001$, suggesting that the dependent samples *t*-test is appropriate in this case. Additionally, paired samples t-tests revealed non-significant differences when the correct, $t(56) = -0.48$, $p = .63$, and incorrect location, $t(56)$

= -0.83, p = .41 was on the left or right-side. Thus, the data was collapsed together when the correct and incorrect location was the left or right-side box. As hypothesized, there was a significantly greater proportion of dwell time to the correct (M = .80, SD = .23) compared to the incorrect (M = .18, SD = .21) location, when a character had a TB, $t(56) = 10.95$, $p < .001$, d = 1.45. A graphical representation of the means and adjusted 99% confidence intervals is shown in Figure 8.

Hypothesis 3 (Implicit ToM task): μ (FB Proportion of First Looks Correct) > μ (FB Proportion of First Looks Incorrect)

A two-way Chi square test for Homogeneity of Proportions was conducted to compare proportion of first looks to the correct location to a hypothetical population in which there was an equal proportion of first looks to the correct and incorrect location. Proportions of first looks to the correct and incorrect location in the FB condition were calculated as follows: First looks to the correct location divided by first looks to the correct and incorrect location equaled proportion of first looks to the correct location; first looks to the incorrect location divided by first looks to the correct and incorrect location equaled proportion of first looks to the incorrect location. Recall that it was randomized whether the correct and incorrect location was on the left or right-side box. The data was collapsed whether the correct and incorrect location was the left or right-side box, as a Chi square test for Homogeneity of Proportions revealed an equal distribution in the population in the FB condition χ^2 (1, N = 30) = 0.08, $p < .90$. The proportion of first looks to the correct (M = .06, SD = .08) and incorrect location (M = .97, SD = .20) was not equal in the population χ^2 (1, N = 30) = 47.73, $p < .01$, $\Phi = 1.26$. Contrary to the hypothesis, there was a significantly greater proportion of first looks to the incorrect location when the character had a FB. This finding mirrored the dwell time data.

Hypothesis 4 (Implicit ToM task): μ (TB Proportion of First Looks Correct) > μ (TB Proportion of First Looks Incorrect)

Hypothesis 4 was statistically analysed in the previously mentioned two-way Chi square test for Homogeneity of Proportions. Proportions of first looks to the correct and incorrect location in the TB condition were calculated the same way as the previous analysis. As in Hypothesis 3, the data was collapsed whether the correct and incorrect location was the left or right-side box, as a Chi square test for Homogeneity of Proportions revealed an equal distribution in the population in the TB condition, χ^2 (1, N = 30) = 0.35, $p < .90$. The proportion of first looks to the correct ($M = .98$, $SD = .22$) and incorrect location ($M = .05$, $SD = .10$) was not equal in the population, χ^2 (1, $N = 30$) = 47.73, $p < .01$, $\Phi = 1.26$. As hypothesized, there was a significantly greater proportion of first looks to the correct location when the character had a TB. This finding mirrored the dwell time data.

H5 (Explicit ToM task): μ (ToM error in Search Question, in FB condition) > μ (ToM error in Memory Question, in FB condition) = μ (ToM error in Search Question, in TB condition) = μ (ToM error in Memory Question, in TB condition)

A 2 x 2 within-subjects ANOVA was conducted to evaluate the effects of belief type (false, true) and question type (search, memory) on ToM error scores; higher scores indicate more ToM errors. Prior to the analysis, Shapiro-Wilk's test was conducted and indicated a violation of the normality assumption for each group (all *p*s ≤ .001). Parametric tests were still conducted because these tests are robust even when there are normality assumption violations. The means, standard deviations, sample size, and 99% confidence intervals for ToM error scores as a function of the two factors are shown in Table 1. The ANOVA indicated a significant main effect for belief type, $F(1, 67) = 68.84$, $p < .001$, partial $\eta^2 = .51$, a significant main effect for question type, $F(1, 67) = 172.58$, $p < .001$, partial $\eta^2 = .72$, and a significant interaction between belief and question type, $F(1, 67) = 74.37$, $p < .001$, partial $\eta^2 = .53$. Because the interaction between belief and question type was significant, the two main effects were ignored, and the question type simple main effects were examined instead. That is, the differences among question type for the FB and TB condition, separately. Contrary to the hypothesis, there were nonsignificant differences between question type conditions for the FB condition, $F(1, 67) = 4.19$, $p < .05$, partial $\eta^2 = .06$, and significant differences for the TB condition, $F(1, 67) = 145.35$, $p < .001$, partial $\eta^2 = .68$.

A follow-up Bonferroni contrast *t*-test was conducted to evaluate the pairwise difference between the search and memory question mean in the TB condition. Contrary to the hypothesis, participants made significantly more ToM errors in the search versus memory question, $p < .001$ ($d = 12.06$).

Exploratory Analyses

In the Explicit ToM task, it was expected that there would be a greater proportion of dwell time towards the incorrect location when participants answered the search question in the FB condition. In comparison, it was expected that there would be a lesser proportion of dwell time towards the incorrect location when participants answered the search and memory question in the TB condition, as well as the memory question in the FB condition. This expectation was supported by previous research on ToM errors that used explicit change-of-location tasks (e.g., Rubio-Fernandez & Glucksberg, 2012). Research on ToM errors informed this expectation because the Explicit ToM task was an explicit change-of-location task; in the Explicit ToM task, participants answered questions about characters' behaviours (Clements & Perner, 1994).

As an exploratory analysis, a 2 x 2 within-subjects ANOVA was conducted to evaluate the effects of belief type (false, true) and question type (search, memory) on the proportion of dwell time to the incorrect location; higher scores indicate greater proportion of dwell time towards the incorrect location. Proportion of dwell time towards the incorrect location was calculated as follows: Dwell time towards the incorrect location divided by dwell time towards the correct and incorrect location equaled proportion of dwell time towards the incorrect location. Prior to the analysis, Shapiro-Wilk's test was conducted and indicated a violation of the normality assumption for all the groups (all *p*s $< .002$) except in the TB condition, when participants answered the search question ($p = .18$). Parametric tests were still conducted because these tests are robust even when there are normality assumption violations. The means, standard deviations, sample size, and 99% confidence intervals for dwell time as a function of the two factors are shown in Table 2. The ANOVA indicated a significant main effect for belief type, $F(1, 44) = 56.59$, $p < .001$, partial $\eta^2 = .56$, a significant main effect for question type, $F(1, 44) = 105.36$, $p < .001$, partial $\eta^2 = .71$, and a significant interaction between belief and question type, $F(1, 44) = 35.22$, $p < .001$, partial $\eta^2 = .45$. Because the interaction between belief and question type was significant, the main effects were ignored, and the question type simple main effects were examined instead. That is, the differences among question type for the FB and TB condition, separately. There were significant differences between question type conditions for the FB condition, $F(1, 44) = 11.84$, $p = .001$, partial $\eta^2 = .21$, and for the TB condition, $F(1, 44) = 78.48$, $p < .001$, partial $\eta^2 = .64$.

A follow-up Bonferroni contrast *t*-test was conducted to evaluate the pairwise difference between the search and memory question mean in the FB condition. Mirroring the dwell time and first look data for the Implicit ToM task, participants' dwell time towards the incorrect location was significantly greater in the search versus memory question in the Explicit ToM task, $p = .001$ ($d = 3.38$); this finding was expected. Similarly, a follow-up Bonferroni contrast *t*-test was conducted to evaluate the pairwise difference between the search and memory question mean in the TB condition. Differing from the dwell time and first look data for the Implicit ToM task, participants' dwell time towards the incorrect location was significantly greater in the search versus memory question in the Explicit ToM task, $p < .001$ ($d = 8.85$); this finding was contrary to my expectation.

As mentioned previously, the statistical analyses for the first four hypotheses were repeated. These statistical analyses were re-run with participants whose report, in the funneled debriefing procedure, indicated that they were unaware or aware of the character's beliefs in the Implicit ToM task. The statistical analyses were repeated to see whether the results for four hypotheses were different when participants' awareness was accounted for. Specifically, the open-ended answers in the funneled debriefing procedure were coded as *did not* ($n = 32$) follow as well as *partially* ($n = 22$) and *fully* ($n = 16$) followed the characters' beliefs in the Implicit ToM task. Statistical analyses were run on participants including awareness of subject as a between-subjects variable. Of these participants, thirteen were excluded because they did not meet the aforementioned selection criterion. These analyses yielded the same pattern of results as the original set of analyses. Thus, participants' eye movements in the Implicit ToM task were similar whether participants were unaware or aware of the character's beliefs.

As mentioned above, the current study used the Implicit ToM task from Cho and Cohen (2017) to replicate their findings. As an exploratory analysis, participants' dwell time in the Implicit ToM task was re-analyzed using Cho and Cohen's (2017) analysis. In the Implicit ToM task, the empty box (i.e., the box without the animal) was where the character should look for the animal in the FB condition. In the TB condition, however, the character should ignore the empty box when he or she looked for the animal, in favour of the actual box (i.e., the box with the animal). Cho and Cohen (2017) found that adult participants showed a greater preference to look at the empty box in the FB versus the TB condition. As in Cho and Cohen's (2017) analysis, a preferential looking ratio to the empty box was calculated in the FB and TB condition as follows: Looking time (LT) to the empty box was divided by looking time to the empty and actual box (i.e., LT to empty box / (LT to empty box + LT to actual box)). A paired samples *t*-test was conducted to evaluate the preferential looking ratio in the FB and TB condition; higher ratios indicate greater preference to look at the empty box. Differing from Cho and Cohen's (2017) findings, there was no greater preferential looking ratio in the FB ($M = .16$, $SD = .19$)

versus the TB condition (*M* = .18, *SD* = .21), *t*(56) = -.79, *p* = .43, Cohen's *d* = -.10. Thus, the current study failed to replicate Cho and Cohen's (2017) findings.

Discussion

In the Implicit ToM task, it was predicted that participants would look longer and first look at the location where a character with a FB or TB (i.e., absent or present when the animal moved) would look for an animal. Contrary to predictions, participants looked longer and first looked at the incorrect location when a character had a FB. As predicted, participants looked longer and first looked at the correct location when a character had a TB. Further, it was predicted that participants would make more ToM errors when answering the search question when the character had a FB. In contrast, it was predicted that participants would make fewer ToM errors when answering the search question when the character had a TB, as well as the memory question in the FB and TB condition (i.e., character was absent or present when the object was moved). Contrary to my prediction, participants made more ToM errors when answering the search question in the TB versus the FB condition, and the memory question in the FB and TB condition.

The current study only partially replicated previous research that used implicit change-of-location tasks (Cho & Cohen, 2017). Previous research found that adult participants looked longer and/or first looked at the correct versus incorrect location in the FB and TB condition (Low & Watts, 2013; Schneider, Bayliss, Becker, & Dux, 2012; Schneider, Lam, et al., 2012; Schneider et al., 2013; Schneider et al., 2014; Schuwerk et al., 2015; Senju et al., 2009).

Further, according to an exploratory analysis, the current study failed to replicated Cho and Cohen's (2017) finding. That is, participants showed no preference for looking at the empty box in the FB compared to the TB condition, when their dwell time in the Implicit ToM task was re-analyzed using Cho and Cohen's (2017) analysis. In other words, participants preferred to look at the empty box about the same amount of time whether the character believed that the animal was (FB condition) or was not (TB condition) in there. If this finding was replicated in future research, it could suggest that participants did not implicitly track the character's beliefs and look, in anticipation, to where the character should look for the animal. Alternatively, the current study may have failed to replicate Cho and Cohen's (2017) findings because of differences in materials that were used (e.g., the dimensions of the videos were smaller in the current study). Overall, further research is necessary to draw any conclusions about failing to replicate Cho and Cohen's (2017) study when the Implicit ToM task was re-analyzed.

In the funneled debriefing procedure, the current study found that about half the participants' self-reports (54.9%) showed that they were partially or fully aware of the character's belief. On the surface, this finding may explain this study's implicit ToM eye-tracking results. For these participants, perhaps the Implicit ToM task tapped into participants' explicit ToM processes, rather than their implicit ToM processes. If this was the case, participants, who were unaware of the characters' beliefs, should look longer and first look at the correct location when the character had a FB. However, when the same analyses were run with participants' awareness (i.e., unaware or aware of the characters' beliefs) as a between-subjects variable, the results showed the same pattern. That is, participants looked longer and first looked to the incorrect or correct location when the character had a FB or TB, respectively.

In the Explicit ToM task, participants made more ToM errors when answering the search question when a character had a TB versus a FB, as well as when participants answered the memory question in the FB and TB conditions. This finding was inconsistent with my prediction. Further, using similar wording as the current study, previous research on the ToM error found that adult participants made more ToM errors when answering the search versus the memory question when a character had a FB (e.g., Sommerville et al., 2013). Although there were larger mean ToM error scores in the search versus memory question in the FB condition of the current study, this difference was not significant with the stringent $p < .01$ criterion that was used. The current study's results may differ from previous research, in part, because different versions of the Explicit ToM task were used. As mentioned before, previous research used the real objects version of the Explicit ToM task, while the current study used a computerized version.

Given the current study's findings, participants may have developed a strategy of clicking on the first location that appeared in each trial to answer the questions, regardless of the belief and question type of the trial. In particular, the first location was the correct answer for the search question in the FB condition, and the memory question in the FB and TB condition. Therefore, the first location would only be the incorrect answer for the search question in the TB condition. Participants may have also been confused by the wording of the scenarios. In particular, one sentence in the scenario was very similar in its wording and word length; however, this sentence determined whether the character had a FB or TB in that trial. Specifically, participants heard "While Dory is *gone*, Clyde moves the fishing rod here," in the FB trial (see Figure 4), and "When Dory *returns*, Clyde moves the fishing rod here," in the TB trial (see Figure 5). Therefore, participants may not have realized that the wording was different and/or that the slight change in wording had implications for the character's beliefs; these misunderstandings would influence participants' answers to the search and memory questions. Thus, it was possible that a confused participant did not explicitly follow, or gave up attempting to explicitly follow, the character's belief. Future research could use a modified funneled debriefing procedure (Schneider, Bayliss, Becker, & Dux, 2012) to check if participants were explicitly following a character's belief, and to inquire about the strategies they used in the task. Additionally, future research should adapt the Explicit ToM task to address participants' strategy of clicking on the first location and their potential confusion.

In the Explicit ToM task, an exploratory analysis found that participants looked longer towards the incorrect location in the search versus the memory question in the FB and TB condition. This finding was expected for the FB, but not for the TB condition. The pattern of results in the FB condition was consistent and TB condition was inconsistent with research on ToM errors, respectively (e.g., Keysar et al., 2000, Rubio-Fernandez & Glucksberg, 2012). In an explicit change-of-location task, Rubio-Fernandez and Glucksberg's (2012) study found that monolingual versus bilingual adults first looked at the incorrect location in the FB trial. Most of the participants, however, first looked at the correct location in the TB trial. Future research could analyze if there was a similar difference between the monolingual and bilingual adults in the current sample. In Keysar et al.'s (2000) study, participants indicated that they thought another person was referring to an object that was only visible to themselves by looking longer at this object than the object that was mutually visible to themselves and the other person. The current study's findings for the TB condition were inconsistent with Rubio-Fernandez and Glucksberg's (2012) study and may also be explained by participants' strategy or confusion in the task.

Limitations and Future Directions

The current study had a sample of predominately undergraduate and female participants, as is commonly found in psychological research. Therefore, future research could use the Implicit ToM and Explicit ToM task, with eye tracking, to study other populations such as children, adolescents, older adults, or even people in special populations (e.g., people with ASD). It was a limitation that different combinations of participants were used for the different statistical analyses; however, it is unrealistic to expect no missing data. Additionally, because the current study only partially replicated Cho and Cohen's (2017) study (and failed to replicate, when the Implicit ToM task was re-analyzed), the proportion of dwell time results in the Explicit ToM task should be treated with caution. Further, it is hard to determine whether the current study's results support implicit ToM processes, as 32 of the 73 participants did not report following a character's belief, and 79.9% of all the participants had an accuracy performance of over 90% on the filler task. There is no clear standard to assess the accuracy of participants' performance, but in Schneider, Bayliss, Becker, and Dux's (2012) study, all their participants had an accuracy performance of over 90% on a different filler task. Participants may not have answered the funneled debriefing procedure in as much detail as they could have because they wanted to finish the study quickly to receive their bonus credit or 10 dollars. Therefore, the results obtained from participants who were unaware of the characters' beliefs should be interpreted with caution. In the future, there should be more incentive for participants to answer these questions completely and honestly. Further, in the current study, the author was the only coder for the eye tracking data in the Implicit ToM and Explicit ToM task as well as the answers in the funneled debrief procedure. Therefore, the author may have shown an experimenter bias. Further research should have another person code this data to have inter-rater reliability. Additionally, to clarify the distinction between implicit and explicit ToM processes, future research can study anticipatory looking by itself or alongside explicit verbal or written predictions of others' behaviour with different populations. Specifically, future research could use an implicit and explicit version of the Sandbox task as well as use an implicit and explicit version of the ToM task.

Conclusion

Overall, the current study showed that adults' ability to understand others' beliefs can be biased by their own private knowledge (e.g., Sommerville et al., 2013). While the current study had mixed results, it did highlight the complexities of studying ToM processes. This study serves as a bridge between different areas of research which involve implicit-discrete-change-of-location tasks and explicit-continuous-change-of-location tasks. In a broad sense, this research contributes to the literature on what people's eye movements can reveal about their ability to understand others' beliefs.

Figures and Tables

Figure 1. The relative size of the scenarios in the Implicit ToM task is shown here.

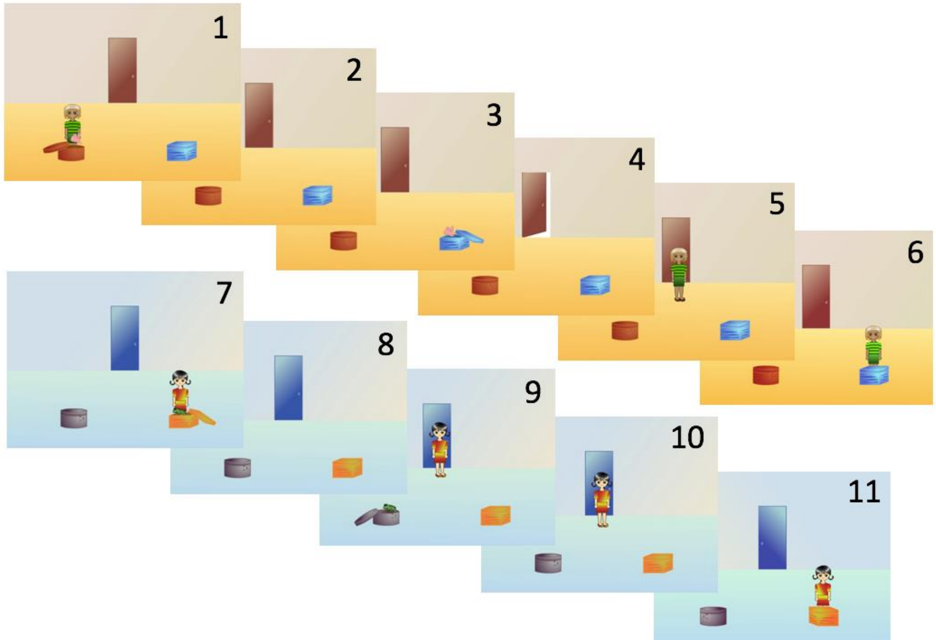

Figure 2. In the Implicit ToM task, participants watched videos. In the FB condition, (1) the character put an animal in one box, (2) the character left, and (3) the animal moved into the other box. Then, (4) the character entered the room, and (5) was about to move towards one box to retrieve the animal. In the TB condition, (7) the character put an animal in one box, (8) exited and (9) entered the room. Then the character saw the animal move into the other box and (10) was about to move towards one box to retrieve the animal. In the FB and TB condition, participants performed a filler task, where they pressed the right (6, 11) or left arrow key to indicate if the character stood behind the right or left-hand box, respectively.

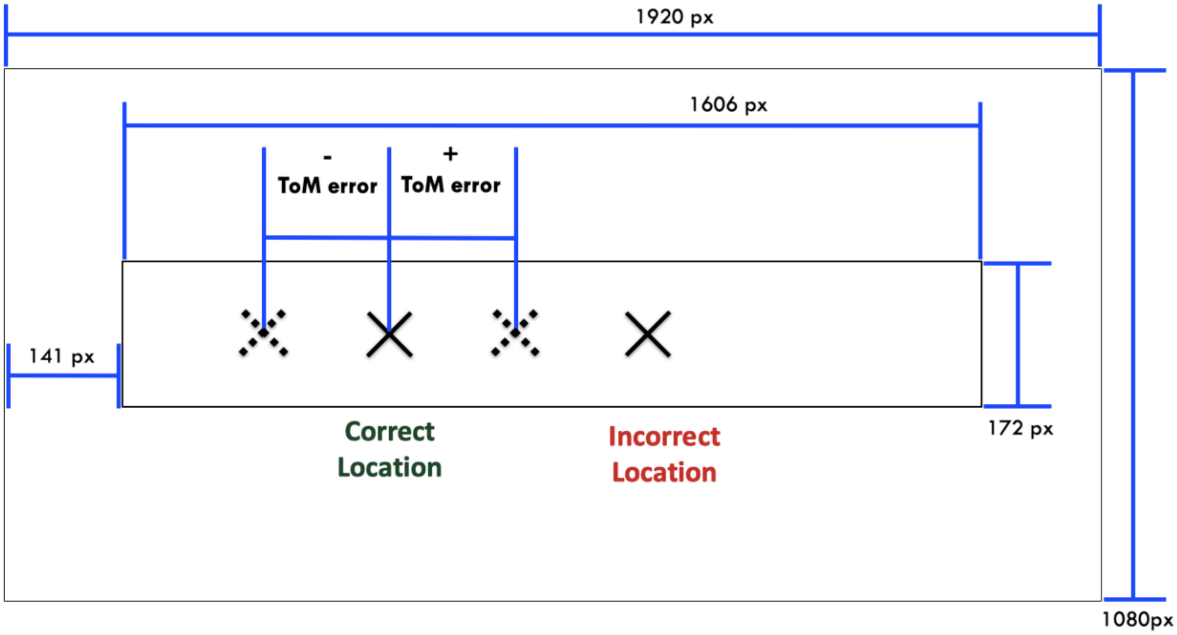

Figure 3. The dimensions of the Explicit ToM task as well as the calculations of ToM error. The dotted 'X's represent participants' response location, or the location that participants indicate to answer the search or memory question. The correct location is the location that was the correct answer to the question. ToM error is the distance (in pixels) between the response location and the correct location. There is a positive ToM error when the response location is towards the incorrect location. However, there is a negative ToM error when the response location is away from the incorrect location.

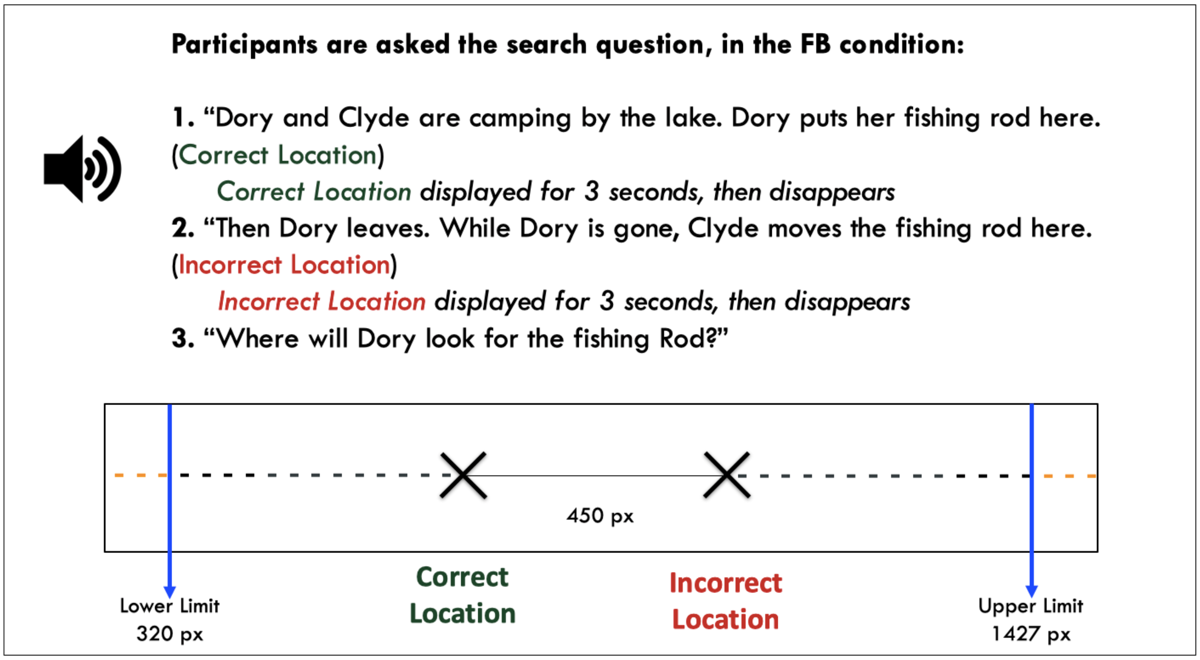

Figure 4. This is the trial design and audio when participants hear the following narrative, in the FB condition. The main characters (1) put an object in one location and (2) leave. In their absence, the secondary characters move the object to another location. Then (3) participants are asked the search question. Participants should indicate that the main characters will look for the object where they initially put it (i.e., the correct location) because they did not see the secondary characters move the object (i.e., the incorrect location). When participants heard the word "here," an "X" appeared in the box to indicate the correct and incorrect location. The correct and incorrect locations are semi-randomly generated so that these locations could appear anywhere within the fixed limits shown (anywhere between the upper and lower limit). Half the trials have the correct location on the right, and the incorrect location on the left. The remaining trials have the correct location on the left, and the incorrect location on the right. There is always 450 pixels between the correct and incorrect location.

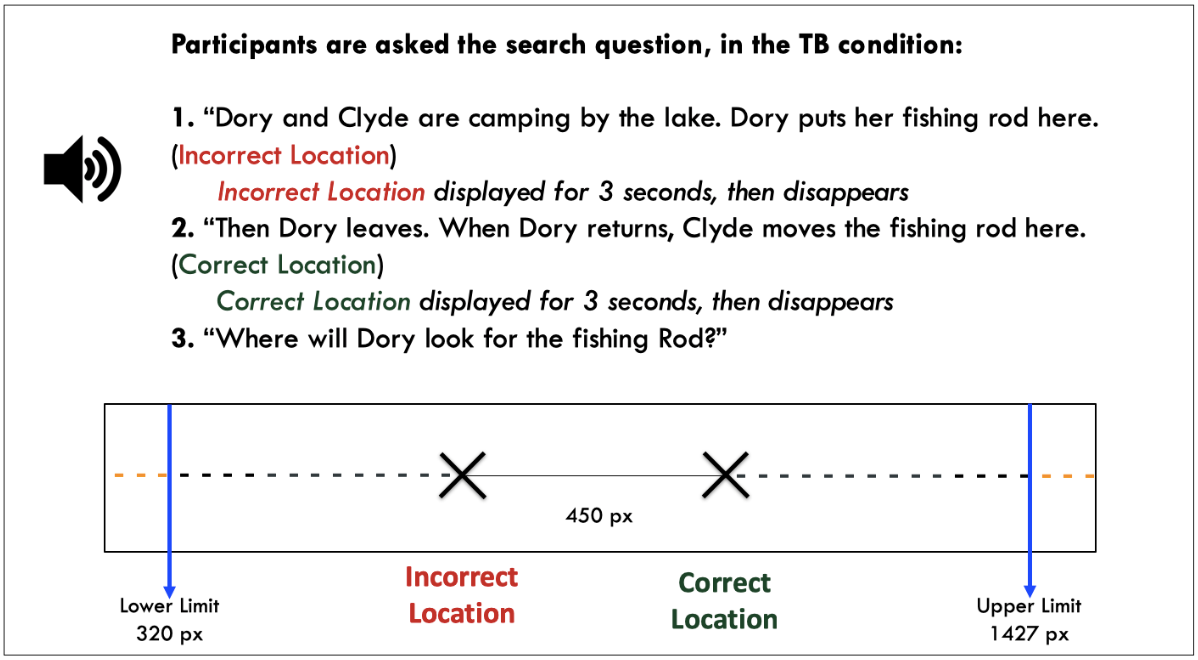

Figure 5. This is the trial design and audio for when participants hear the following narrative, in the TB condition. The main characters (1) put an object in one location, (2) leave and return to the scene. In their presence, the secondary characters move the object to another location. Then (3) participants are asked the search question. Participants should indicate that the main characters will look for the object where they saw the secondary characters move the object to (i.e., the correct location). When participants heard the word "here," an "X" appeared in the box to indicate the correct and incorrect location. The correct and incorrect locations could appear anywhere within the fixed limits shown (anywhere between the upper and lower limit).

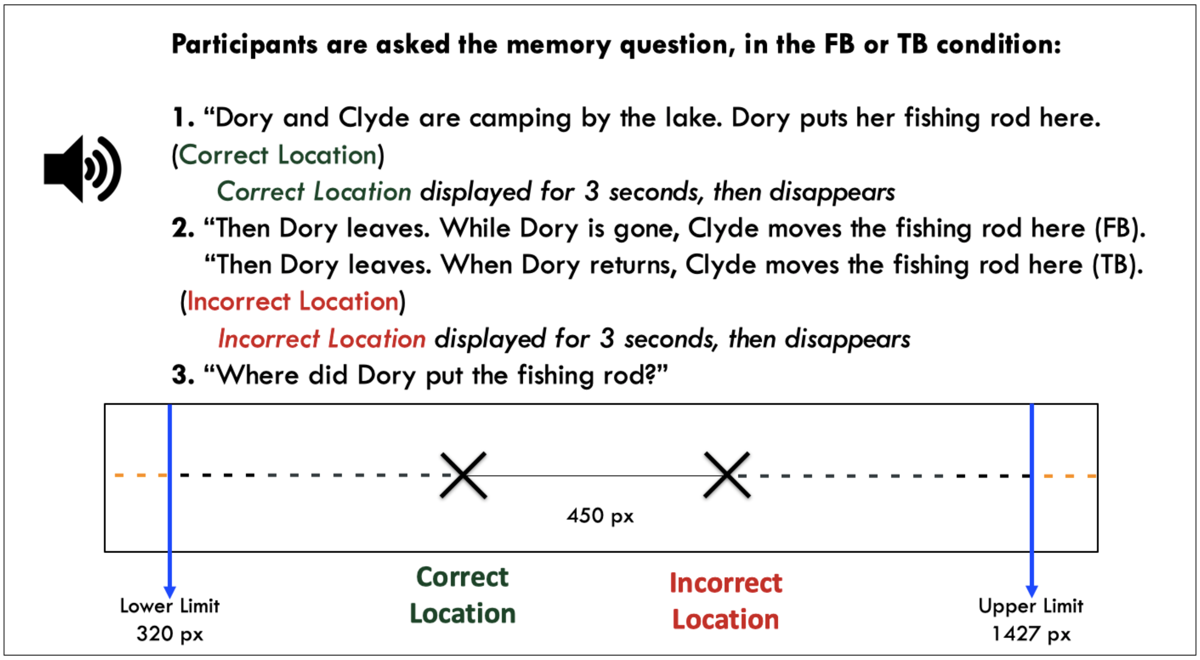

Figure 6. This is the trial design and audio for when participants hear the following narrative, in the FB or TB condition. The main characters (1) put an object in one location and (2) leave, or leave and return to the scene, respectively. In their absence or presence, the secondary characters move the object to another location. Then (3) participants are asked the memory question, and they should indicate the location where the main characters put the initially object (i.e., the correct location). When participants heard the word "here," an "X" appeared in the box to indicate the correct and incorrect location. The correct and incorrect locations could appear anywhere within the fixed limits shown (anywhere between the upper and lower limit).

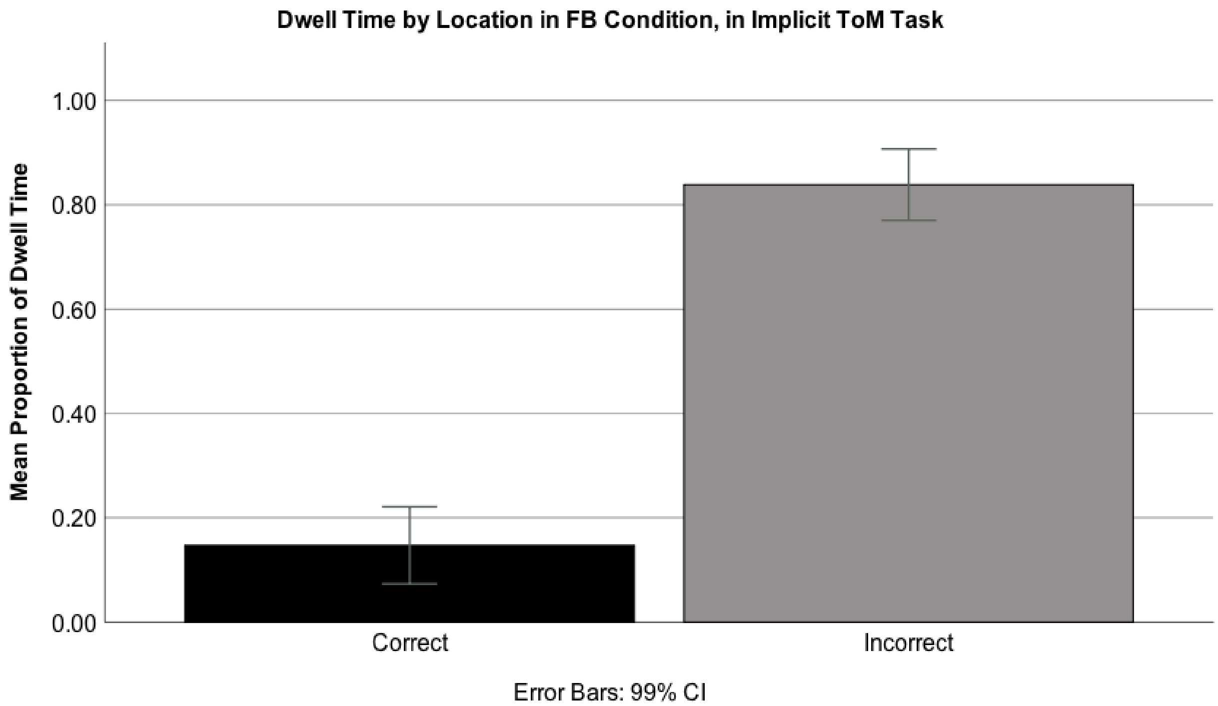

Figure 7. The means and adjusted 99% confidence intervals appear here. Bars represent the mean proportion of dwell time towards the correct and incorrect location, and error bars show adjusted 99% confidence intervals. When a character had a FB (i.e., absent when the animal moved), there was a significantly greater proportion of dwell time to the incorrect compared to the correct location; this finding was contrary to the hypothesis.

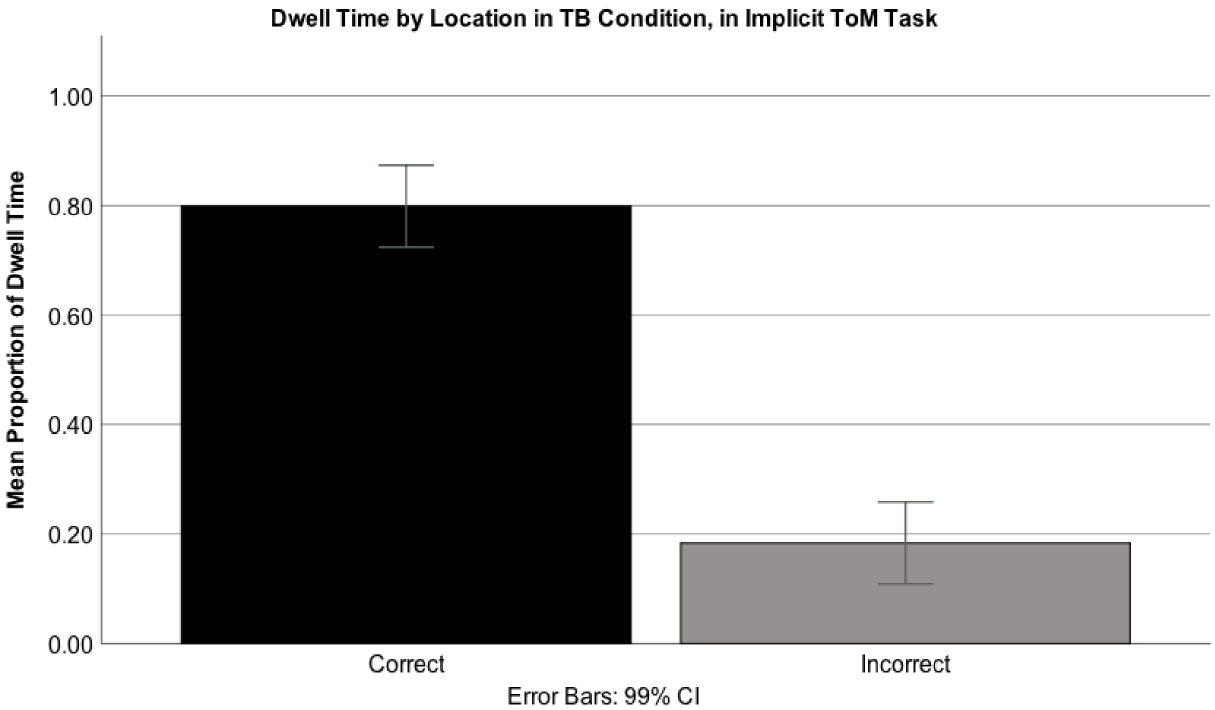

Figure 8. The means and adjusted 99% confidence intervals appear here. Bars represent the mean proportion of dwell time towards the correct and incorrect location, and error bars show adjusted 99% confidence intervals. When a character had a TB (i.e., present when the animal moved), there was a significantly greater proportion of dwell time to the correct compared to the incorrect location; this finding was hypothesized.

Table 1

Means, Standard Deviations, Sample Size, and 99% Confidence Intervals for ToM Error Scores in Explicit ToM Task

		ToM Error Scores		
Belief Type	Question Type	*M*	*SD*	99% CI
False Belief	Search Question	35.49	102.95	[2.39, 68.58]
	Memory Question	11.00	50.13	[-5.12, 27.12]
True Belief	Search Question	267.36[a]	162.96	[214.96, 319.75]
	Memory Question	6.69[a]	48.97	[-9.05, 22.44]

Note. n = 68 for each group. CI = confidence interval. Results of current study. Average ToM error scores when participants answered the search question (Where will the character look for the object?) and memory question (Where did the character put the object?). The character had either a FB (i.e., absent when the object was moved) or a TB (i.e., present when the object was moved). Higher scores indicate more ToM errors. Within each belief type, means in the same column sharing superscripts differ at $p < .001$ in the pairwise comparisons, with a Bonferroni correction.

Table 2

Means, Standard Deviations, Sample Size, and 99% Confidence Intervals for Proportion of Dwell Time towards Incorrect Location in Explicit ToM Task

		Proportion of Dwell Time Towards Incorrect Location		
Belief Type	Question Type	*M*	*SD*	99% CI
False Belief	Search Question	.21[a]	.19	[.14, .29]
	Memory Question	.13[a]	.13	[.08, .18]
True Belief	Search Question	.61[b]	.25	[.51, .72]
	Memory Question	.20[b]	.16	[.14, .26]

Note. $n = 68$ for each group. CI = confidence interval. Results of current study. Average proportion of dwell time towards incorrect location when participants answered the search question (Where will the character look for the object?) and memory question (Where did the character put the object?). The character had either a FB (i.e., absent when the object was moved) or a TB (i.e., present when the object was moved). Higher scores indicate greater proportion of dwell time towards the incorrect location. Within each belief type, means in the same column sharing superscripts differ at $p < .01$ in the pairwise comparisons, with a Bonferroni correction.

References

Apperly, I. A., & Butterfill, S. (2009). Do humans have two systems to track beliefs and belief-like states?. *Psychological Review, 116*, 953–970. doi:10.1037/a0016923

Apperly, I. A., Riggs, K., Simpson, A., Chiavarino, C., & Samson, D. (2006). Is belief reasoning automatic?. *Psychological Science, 17*, 841–844. doi:10.1111/j.1467-9280.2006.01791.x

Baddeley, A. D. (1992). Working memory. *Science*, *255*, 556–559.

Bargh, J. A. (1994). The four horsemen of automaticity: Awareness, intention, efficiency, and control in social cognition. In R. S. Wyer Jr. & T. K. Srull (Eds.), *Handbook of social cognition* (2nd ed., pp. 1–40). Hillsdale, NJ: Erlbaum.

Bargh, J. A., & Chartrand, T. L. (2000). The mind in the middle: A practical guide to priming and automaticity research. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 253–285). New York: Cambridge University Press.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a “theory of mind”?. *Cognition*, *21*, 37–46. https://doi.org/10.1016/0010-0277(85)90022-8

Begeer, S., Bernstein, D. M., van Wijhe, J., Scheeren, A. M., & Koot, H. M. (2012). A continuous false belief task reveals egocentric biases in children and adolescents with autism spectrum disorders. *Autism*, *16*, 357–366. doi:10.1177/1362361311434545

Bernstein, D. M., Coolin, A., Fischer, A. L., Thornton, W. L., & Sommerville, J. A. (2017). False-belief reasoning from 3 to 92 years of age. *PLoS One, 12*(9), e0185345. doi:10.1371/journal.pone.0185345

Bernstein, D. M., Thornton, W. L., & Sommerville, J. A. (2011). Theory of mind through the ages: Older and middle-aged adults exhibit more errors than do younger adults on a continuous false belief task. *Experimental Aging Research, 37*(5), 481-502. doi:10.1080/0361073X.2011.619466

Clements, W. A., & Perner, J. (1994). Implicit understanding of belief. *Cognitive Development*, *9*, 377–395. https://doi.org/10.1016/0885-2014(94)90012-4

Cho, I. & Cohen, A.S. (2017). Implicit Theory of mind in older adults: Are there two systems for mindreading? Poster presented at: Cognitive Development Society Bi-Ennial Conference 2017; 2017 Oct 12-14; Portland, OR.

Coburn, P. I., Bernstein, D. M., & Begeer, S. (2015). A new paper and pencil task reveals adult false belief reasoning bias. *Psychological Research*, *79*, 739–749. doi:10.1007/s00426-014-0606-0

Cohen, A. S., & German, T. C. (2009). Encoding of others' beliefs without overt instruction. *Cognition*, *111*(3), 356-363. doi:10.1016/j.cognition.2009.03.004

Cohen, A. S., & German, T. C. (2010). A reaction time advantage for calculating beliefs over public representations signals domain specificity for 'theory of mind'. *Cognition, 115*, 417–425. doi:10.1016/j.cognition.2010.03.001

Derksen, D. G. (2017). Avoidance-Based Scenarios Inflate Theory-of-Mind Bias in the Sandbox (unpublished honours thesis). Kwantlen Polytechnic University, Surrey, Canada.

Diamond, A. (2013). Executive functions. *Annual Review of Psychology, 64*, 135-168.

Dumontheil, I., Apperly, I. A., & Blakemore, S. J. (2010). Online usage of theory of mind continues to develop in late adolescence. *Developmental Science*, *13*, 331–338. doi:10.1111/j.1467-7687.2009.00888.x

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175-191.

Garnham, W. A., & Ruffman, T. (2001). Doesn't see, doesn't know: is anticipatory looking really related to understanding or belief?. *Developmental Science*, *4*(1), 94-100. doi:10.1111/1467-7687.00153

German, T. C., & Cohen, A. S. (2012). A cue-based approach to "theory of mind": Re-examining the notion of automaticity. *British Journal of Developmental Psychology*, *30*, 45–58. doi: 10.1111/j.2044-835X.2011.02055.x

Giesbrecht, A. L., Bernstein, D. M., & Derksen, D. G. (2018, February 19). What Eye Tracking Reveals in an Implicit-Discrete Versus an Explicit-Continuous Theory-of-Mind Measure. Retrieved from osf.io/ft3ax

Henry, J. D., Phillips, L. H., Ruffman, T., & Bailey, P. E. (2013). A meta-analytic review of age differences in theory of mind. *Psychology and Aging, 28*(3), 826-839. doi:10.1037/a0030677

Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science*, *11*, 32–38. doi:10.1111/1467-9280.00211

Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition*, *89*, 25–41. doi:10.1016/S0010-0277(03)00064-7

Kovács, Á. M., Téglás, E., & Endress, A. D. (2010). The social sense: Susceptibility to others' beliefs in human infants and adults. *Science, 330*, 1830–1834. doi:10.1126/science.1190792

Leslie, A. M., Friedman, O., & German, T. P. (2004). Core mechanisms in "theory of mind". *Trends in Cognitive Sciences, 8*, 528–533. doi:10.1016/j.tics.2004.10.001

Leslie, A. M., & Polizzi, P. (1998). Inhibitory processing in the false belief task: Two conjectures. *Developmental Science, 1*, 247-253. doi:10.1111/1467-7687.00038

Low, J., & Watts, J. (2013). Attributing false beliefs about object identity reveals a signature blind spot in humans' efficient mind-reading system. *Psychological Science*, *24*, 305-311. doi:10.1177/0956797612451469

Mahy, C. E. V., Bernstein, D. M., Gerrard, L. D., & Atance, C. M. (2017). Testing the validity of a continuous false belief task in 3- to 7-year-old children. *Journal of Experimental Child Psychology, 160*, 50-66. doi:10.1016/j.jecp.2017.03.010

Onishi, K., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science*, *308*, 255–258. doi:10.1126/science.1107621

Perner, J., & Ruffman, T. (2005). Infants' insight into the mind: How deep? *Science*, *308*, 214-216. doi:10.1126/science.1111656

Posten, H. O. (1984). Robustness of the two-sample t-test. In D. Rasch and M.L. Tiku (Eds.), *Robustness of statistical methods and nonparametric statistics* (pp. 92-99). Dordrecht, Germany: Reidel.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind?. *Behavioral and Brain Sciences*, *1*(04), 515-526. https://doi.org/10.1017/S0140525X00076512

Rubio-Fernandez, P. (2017). Can we forget what we know in a false-belief task? An investigation of the true-belief default. *Cognitive Science, 41*, 218-241. doi:10.1111/cogs.12331

Rubio-Fernandez, P., & Glucksberg, S. (2012). Reasoning about other people's beliefs: bilinguals have an advantage. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 211-217. doi:10.1037/a0025162

Ruffman, T. (2014). To belief or not belief: Children's theory of mind. *Developmental Review*, *34*, 265-293. doi:10.1016/j.dr.2014.04.001

Schneider, D., Bayliss, A. P., Becker, S. I., & Dux, P. E. (2012). Eye movements reveal sustained implicit processing of others' mental states. *Journal of Experimental Psychology: General*, *141*, 433–438. doi:10.1037/a0025458

Schneider, D., Lam, R., Bayliss, A. P., & Dux, P. E. (2012). Cognitive load disrupts implicit theory-of-mind processing. *Psychological Science, 23,* 842–847. doi:10.1177/0956797612439070

Schneider, D., Nott, Z. E., & Dux, P. E. (2014). Task instructions and implicit theory of mind. *Cognition*, *133,* 43-47. doi:10.1016/j.cognition.2014.05.016

Schneider, D., Slaughter, V. P., Bayliss, A. P., & Dux, P. E. (2013). A temporally sustained implicit theory of mind deficit in autism spectrum disorders. *Cognition*, *129*, 410–417. doi:10.1016/j.cognition.2013.08.004

Schneider, D., Slaughter, V. P., & Dux, P. E. (2015). What do we know about implicit false-belief tracking?. *Psychonomic Bulletin & Review, 22*, 1-12. doi:10.3758/s13423-014-0644-z

Schuwerk, T., Vuori, M., & Sodian, B. (2015). Implicit and explicit theory of mind reasoning in autism spectrum disorders: The impact of experience. *Autism, 19,* 459-468. doi:10.1177/1362361314526004

Senju, A., Southgate, V., White, S., & Frith, U. (2009). Mindblind eyes: An absence of spontaneous theory of mind in Asperger syndrome. *Science, 325,* 883–885. doi:10.1126/science.1176170

Sommerville, J. A., Bernstein, D. M., & Meltzoff, A. N. (2013). Measuring beliefs in centimeters: Private knowledge biases preschoolers' and adults' representation of others' beliefs. *Child Development*, *84,* 1846–1854. doi:10.1111/cdev.12110

Southgate, V., Senju, A., & Csibra, G. (2007). Action anticipation through attribution of false belief by 2-year-olds. *Psychological Science*, *18*, 587–592. doi:10.1111/j.1467-9280.2007.01944.x

Wang, L., & Leslie, A. M. (2016). Is implicit theory of mind the 'Real Deal'? The own-belief/true-belief default in adults and young preschoolers. *Mind & Language*, *31*, 147-176. doi:10.1111/mila.12099

Wellman, H., Cross, D., &Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development*, *72*, 655–684. doi:10.1111/1467-8624.00304

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*, 103–128. https://doi.org/10.1016/0010-0277(83)90004-5

Van Gaal, S., Ridderinkhof, K. R., Fahrenfort, J. J., Scholte, H. S., & Lamme, V. A. F. (2008). Frontal cortex mediates unconsciously triggered inhibitory control. *Journal of Neuroscience*, *28*(32), 8053-8062. doi:10.1523/JNEUROSCI.1278-08.2008

Appendix A

Audio Transcription in Explicit ToM task

This study included the Explicit ToM task, which used 12 trials. Each trial was presented in a random order and in either the FB or TB condition. The events of one trial are outlined below. Finally, there is a table listing the main and secondary characters' names, as well as the objects, and activities included in each trial.

"Dory and Clyde are camping by the lake. Dory puts her fishing rod, here." *There was an 'X' on the screen for 3 seconds which indicated where the main character put an object.* "Then Dory leaves. While Dory is gone, Clyde moves the fishing rod here," (aka., FB) or "Then Dory leaves. When Dory returns, Clyde moves the fishing rod here" (aka., TB). *There was an 'X' on the screen for 3 seconds which indicated where a secondary character put an object.* Afterwards, participants were asked, "Where will Dory look for the fishing Rod?" (aka., Search Question). They were also be asked, "Where did Dory put the fishing rod? (aka., Memory Question). *In response to both questions, participants used a computer mouse to click on a location, that was within a box on the screen.*

Trial #	Main Character	Secondary Character	Object	Activity
2	Janice	Neetu	chalk	playing hopscotch
3	Jason	Billy	popcorn	watching a movie
4	Maurice	Sam	spoon	eating cake
5	Marian	Shahira	teapot	drinking tea
6	Mark	Tony	hammer	fixing the shed
7	Sally	Ann	toy	playing in the sandbox
8	Yoko	Katelyn	ice cream	putting the groceries away
9	Jenny	Carissa	backpack	walking in the forest
10	Julie	Andrea	socks	putting away clothes
11	Eric	Teuta	paint brush	painting
12	Karan	Jasper	camera	taking pictures

Appendix B

Funneled Debriefing Procedure

This funneled debriefing procedure has been modified from Schneider, Bayliss, Becker, and Dux's (2012) study, who adapted this procedure from Bargh and Chartrand (2000). Participants responded to each question on the computer.

1. What do you think the purpose of the experiment was?
2. What do you think this experiment was trying to study?
3. Did you think that any of the tasks you did were related in any way? (If "yes") In what way were they related?
4. Did anything you did on one task affect what you did on any other task? (If "yes") How exactly did it affect you?
5. When you were completing the task to detect where a character is standing, did you notice anything unusual about the movies?
6. Did you notice any particular pattern or theme to the movies that were included in the task to detect where a character is standing?
7. What were you trying to do while watching the movies and complete the task to detect where a character is standing? Did you have any particular goal or strategy?