A new paper and pencil task reveals adult false belief reasoning bias

Patricia I. Coburn · Daniel M. Bernstein ·
Sander Begeer

Abstract Theory of mind (ToM) is the ability to take other people's perspective by inferring their mental state. Most 6-year olds pass the change-of-location false belief task that is commonly used to assess ToM. However, the change-of-location task is not suitable for individuals over 5 years of age, due to its discrete response options. In two experiments, we used a paper and pencil version of a modified change-of-location task (the Real Object Sandbox task) to assess false belief reasoning continuously rather than discretely in adults. Participants heard nine change-of-location scenarios and answered a critical question after each. The memory control questions only required the participant to remember the object's original location, whereas the false belief questions required participants to take the perspective of the protagonist. Participants were more accurate on memory trials than trials requiring perspective taking, and performance on paper and pencil trials correlated with corresponding trials on the Real Object Sandbox task. The Paper and Pencil Sandbox task is a convenient continuous measure of ToM that could be administered to a wide range of age groups.

P. I. Coburn (✉) · D. M. Bernstein
Kwantlen Polytechnic University, Surrey, BC, Canada
e-mail: tcoburn@sfu.ca

Present Address:
P. I. Coburn
Simon Fraser University, Burnaby, BC, Canada

S. Begeer
VU University, Amsterdam, The Netherlands

S. Begeer
University of Sydney, Sydney, Australia

Introduction

Perspective taking is essential to social interaction, and requires individuals to reason about others' mental states including their beliefs, emotions and intentions (Nickerson, 1999). Together these skills comprise theory of mind (ToM). Researchers often examine ToM by evaluating the understanding of mental states, most notably false beliefs, in young, typically developing children or in children with autism (Yirmiya, Erel, Shaked, & Solomonica-Levi, 1998; Wellman, Cross, & Watson, 2001). The most commonly used measure assesses first-order false belief reasoning, a component of ToM, with a change-of-location task (Wimmer & Perner, 1983; Baron-Cohen, Leslie, & Frith, 1985). However, this task is most appropriate for pre-schoolers and shows ceiling effects when administered to older children and adults. The current study uses a modified change-of-location task to reveal false belief reasoning errors in young adults.

False belief reasoning requires individuals to make judgements about another person's behavior when that person has a false belief about a situation. The classic change-of-location task measures first-order false belief reasoning—understanding the thoughts and intentions of a person who holds a false belief—as a discrete variable, based on children's responses to hypothetical scenarios about social interactions (Wimmer & Perner, 1983). To pass this task, individuals must take the perspective of a protagonist who is unaware that an object has moved from one location to a second location. One scenario depicts Sally and Anne playing with a ball. Sally places the ball in the cupboard and then leaves the room. While Sally is gone, Anne moves the ball from the cupboard to a box. Participants must decide where Sally will search for the ball upon returning, and therefore must ignore their own

perspective and understand that Sally has a false belief about where the ball actually is. Participants who ignore their privileged knowledge are deemed to have a mature ToM, while those who fail to ignore their privileged knowledge are deemed to have an immature ToM.

Most children under 4 years of age fail the false belief task (although see Perner & Roessler, 2012). By age 6, the majority of typically developing children pass this task, leading to the assumption that children older than age 6 have developed a first-order ToM understanding (Wellman et al., 2001). However, few studies have tested the assumption that children older than six have first-order ToM using age-appropriate measures for first-order ToM (Miller, 2009; although see Carpendale & Chandler, 1996). Interestingly, the studies that use appropriate measures to focus on first-order ToM in normal adults generally reveal perspective-taking errors: Adults show an egocentric bias (a tendency to err according to privileged knowledge) in their responses, indicating a failure in first-order ToM (Keysar, Lin, & Barr, 2003; Surtees & Apperly, 2012).

In one experiment, participants (aged 8, 10, and 21 years) viewed an avatar who stood facing one direction in a cartoon room, resulting in only three walls being visible to the avatar. The participants' task was to take a self-perspective or the avatar's perspective and determine how many dots along the walls of the room were visible (Surtees & Apperly, 2012). Participants heard an auditory recording stating that either they or the avatar could see N dots, where N was one to three dots, and were asked to indicate whether this was correct. All age groups showed egocentric errors by answering more accurately from their own perspective than from the avatar's perspective.

Similar findings emerged from a communication experiment in which one participant (the director) instructed another participant (the addressee) to move objects around a 4 × 4 array of slots, containing individual objects in each slot (Keysar, Lin, & Barr, 2003). The objects in some of the slots were only visible to the addressee, and hidden from the director. To determine which object the director was referring to, the addressee had to employ ToM and assume the director's perspective. In the experimental condition, the hidden object had an ambiguous name, similar to the target object that was to be moved. For example, when asked to relocate 'the large cup', in one slot there was a small cup and in another slot there was a larger cup, both of which were visible to the participant and the director. However, in the hidden slot there was an even larger cup. Although participants were aware that the director could not see the hidden object, surprisingly often they tried to move the hidden object when it fit the description of the to-be-moved object. Thus, participants were unable to ignore their privileged perspective when attempting to take the director's naïve perspective. These findings stand in stark contrast to the mature false belief reasoning of children aged 6 and older who pass the classic ToM tasks. This discrepancy is likely due to the varying methods and tasks researchers use to assess ToM in different age groups. Having a task that could easily be administered to all age groups without procedural changes would add to the emerging literature on the development of perspective taking across the lifespan.

In order to develop a first-order false belief task for a wider age range, Sommerville, Bernstein, and Meltzoff (2013) adapted the classic change-of-location task to measure first-order false belief in preschoolers and adults using a continuous outcome scale. This new task, which we refer to here as the Real Object Sandbox task, uses multiple false belief and memory control trials and a virtually unlimited number of response options capable of detecting false belief on a continuum. In the task, participants face the researcher with a large Styrofoam-filled sandbox between them. The researcher reads a scenario involving a protagonist who places an object in one location within the box, and then leaves the room. Another character then moves the object to a second location within the box. To illustrate these scenarios, the researcher locates and relocates real objects within the Styrofoam contained in the sandbox. The researcher asks the participant either a critical false belief or memory control question. The false belief question requires the individual to set aside his/her own privileged knowledge and take the naïve perspective of the protagonist, while the memory control question requires only that the individual remember the original location of the object. Compared to the classic change-of-location task, the Real Object Sandbox task allows for the object to be placed, and more critically, for the individual to respond to a location anywhere along the length of the 5-foot container. This allows for errors, operationalized as bias, to be measured as a continuous rather than a discrete variable.

Recently, we created a more convenient paper and pencil version of the Real Object Sandbox task (Begeer, Bernstein, van Wijhe, Scheeren, & Koot, 2012). We compared performance in typically developing adolescents and adolescents with autism, using a false belief scenario and a no-false belief control scenario, in which the protagonist, respectively, does not or does know the hidden object's current location. In the no-false belief scenario, the protagonist hid a different object in a second location. Therefore, this trial did not involve a change of location of the original object. Both groups showed more bias on the false belief than the no-false belief scenario, while individuals with autism showed significantly more bias than the typically developing adolescents. Perspective-taking deficits in clinical populations have a profound effect on social functioning.

Thus, the detection of these deficits is an important component of assessment.

Rather than using age-appropriate first-order ToM measures, various researchers have attempted to make ToM measures more suitable for older children and adults by increasing task complexity, resulting in advanced ToM measures, which primarily include "second-order" false belief tasks. These tasks assess the understanding of embedded mental states (e.g., thinking about what another person thinks about what a third person thinks; Tager-Flusberg & Sullivan, 1994) and successfully measure ToM development in older children (Perner & Wimmer, 1985). However, typically developing 9-year olds often perform at ceiling on these tasks (Happé, 1994). Advanced ToM measures are generally unsuitable for (pre) school-aged children; moreover, the validity of these measures may be undermined by the fact that they assess related, general cognitive skills such as language and executive function rather than targeting core first-order ToM abilities (Tager-Flusberg & Sullivan, 1994; Scheeren, de Rosnay, Koot, & Begeer, 2013; see also Rakoczy, Harder-Kasten, & Sturm, 2012; Rubio-Fernandez & Geurts, 2012). Finally, individuals employ first-order ToM frequently in everyday interactions, for example when determining what object someone is describing and providing one's conversation partner with adequate information. Conversely, people seem to use second-order ToM skills in specific situations, such as playing complex card games such as poker or bridge. First-order ToM skills are thus more frequently employed than second-order ToM skills during everyday social interactions.

Currently, there are few simple tasks that can be administered to preschoolers, older children and adults to measure false belief reasoning without altering the procedures. Developing a simple measure that could fulfill this requirement would be a significant contribution to ToM research. The paper and pencil version of the Sandbox task could satisfy this requirement if the egocentric perspective-taking results from Begeer et al. (2012), which only examined adolescent performance, can be replicated in other age groups such as adults. Additionally, the Sandbox task version used by Begeer et al. (2012) included only one trial to measure false belief reasoning and one trial to measure no-false belief reasoning. Moreover, the latter condition involved participants' memory for a different object being placed in a second location within the sandbox. It is, therefore, possible that this no-false belief control condition failed to control adequately for perspective taking in the false belief condition.

The purpose of the current study was to administer the Paper and Pencil Sandbox task to assess first-order false belief performance in young adults. In Experiment 1, we included four false belief and four memory control trials derived from the Real Object Sandbox task and measured location accuracy as a bias toward the incorrect location (Sommerville et al., 2013). Comparing performance across several trials of mental state and non-mental state reasoning provides a robust measure of false belief reasoning. Utilizing several memory control trials instead of a single no-false belief trial, as done by Begeer et al. (2012), could illuminate perspective-taking difficulties that may exist in the face of general cognitive abilities such as working memory. Based on prior work, we predicted more bias on the false belief trials than the memory control trials (cf. Begeer et al., 2012; Sommerville et al., 2013; Surtees & Apperly, 2012). In Experiment 2, we sought to validate the Paper and Pencil Sandbox task by comparing it to performance on the Real Object Sandbox task.

Experiment 1

We administered the Paper and Pencil Sandbox task to examine if individuals would show more bias on trials that require perspective taking than on trials that only require them to remember the original location of an object. This would indicate that adults demonstrate false belief reasoning bias.
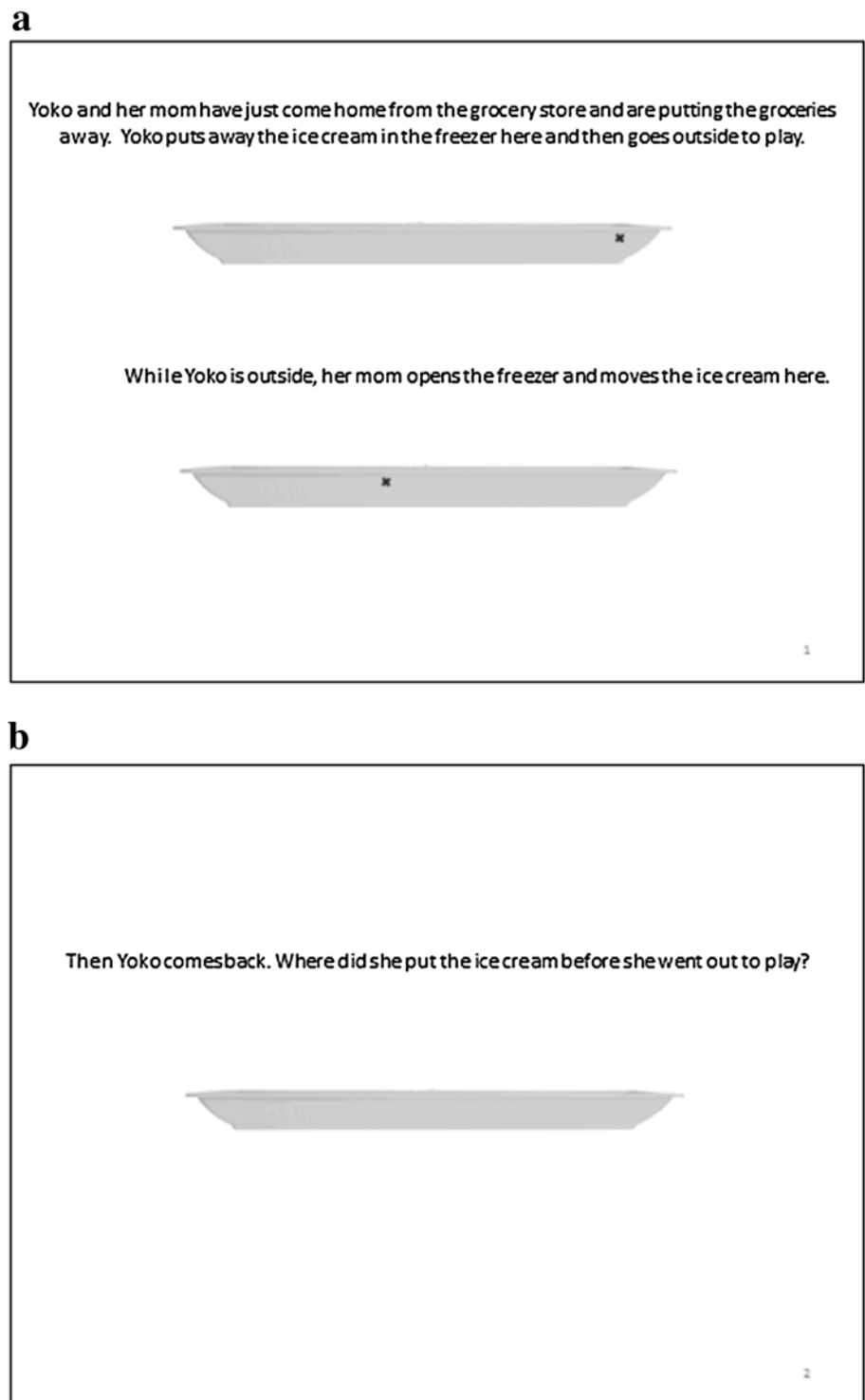
Method

Participants

Eighty-one undergraduates from a Canadian university (M age = 20.80 years; Range 18–29; 68 % female) received course credit through the psychology research pool. Participants completed the Paper and Pencil Sandbox task and then completed an unrelated task as part of another study.

Materials

Paper and Pencil Sandbox task The Paper and Pencil Sandbox task is a modification of the original, Real Object Sandbox task that involved a 5-foot long Styrofoam-filled box (Sommerville et al., 2013; see Begeer et al., 2012). The Paper and Pencil Sandbox task used here involved two versions of a container (e.g., sandbox, freezer, garden plot) printed on one side of an 8.5″ × 11″ piece of paper (see Fig. 1a, b), and a single version of the container printed on the other side of the paper. The Paper and Pencil Sandbox task consisted of nine trials. On each trial, participants heard a different change-of-location scenario and then responded to a critical question. We used a word search book as a brief filler task between the scenario and response portion of each trial.

Fig. 1 a Example of page one—Paper and Pencil Sandbox scenario (wording for memory control and false belief was identical on the first page). **b** Example of flip side of page one—Paper and Pencil Sandbox scenario (memory control trial)

a

Yoko and her mom have just come home from the grocery store and are putting the groceries away. Yoko puts away the ice cream in the freezer here and then goes outside to play.

While Yoko is outside, her mom opens the freezer and moves the ice cream here.

b

Then Yoko comes back. Where did she put the ice cream before she went out to play?

The first side of the paper had an "×" on each container to indicate the original location (L1), in which a story character hides an object, and (L2), the change of location. The second side of the paper had no marking on the container to allow the participant to respond to the critical question. Each container was 148 millimeters long. The participant heard a different scenario on each trial. There were nine trials and therefore nine different scenarios. In each scenario, a protagonist placed an object in the container (L1). While the protagonist was away (false belief and memory control trials), another character moved the object to another location (L2) within the container. On one

trial (true belief), the protagonist watched as the other character moved the object. The four false belief trials required participants to take the protagonist's perspective, who had a false belief about the object's location, whereas the memory control trials only required participants to remember the original location of the objects. We included the true belief trial to prevent participants from realizing that the correct response on every other trial was L1. The correct response to the true belief trial was L2 rather than L1, because the protagonist watches the object being relocated. The trials followed a fixed order. The participants completed two memory control trials, two false belief trials, a true belief trial, two more false belief trials and finally two more memory control trials.

Design

We used a single factor (belief: false belief, memory control) within-subjects design.

Procedure

This study was approved by the Kwantlen Polytechnic University Research Ethics Board. Participants provided full informed consent before completing this study. Each participant heard nine scenarios and completed the task within 10 min. On a false belief trial, the researcher would read the first part of the scenario to the participant. For example, "Yoko and her mom have just come home from the grocery store and are putting the groceries away. Yoko puts the ice cream in the freezer here and then goes outside to play." The researcher then turned the paper to show the participant the container here denoting the freezer with an "X" on it to indicate the location (L1). The paper had both parts of the scenario on it, therefore the researcher used a piece of paper to cover the part that was not meant to be viewed currently. After the participant had the opportunity to see L1 the researcher collected the paper and read the second part of the scenario, covering the first part of the scenario. "While Yoko is outside, her mom opens the freezer and moves the ice cream here." The researcher then told the participant to search for words in a word search book. We administered this distractor task to prevent participants from using perceptual strategies to guide their search in the Sandbox task. After 20 s, the researcher turned the page to read the critical question: "When Yoko comes back, where will she look for the ice cream?" The researcher then showed the paper with another copy of the container on it. This one was blank allowing the participant to write an "×" to indicate where they thought Yoko would look for the ice cream. The correct response to this trial was L1.

On a memory control trial, the researcher would use the same procedures as on the false belief trial; however, the

Table 1 Mean bias (standard deviation) in proportions for the Paper and Pencil Sandbox task and the Real Object Sandbox task (Experiment 1 and Experiment 2)

	Experiment 1		Experiment 2	
	Paper and pencil 70 mm	Paper and pencil 35 mm	Paper and pencil 35 mm	Real object 35 cm
False belief	0.17 (0.19)	0.06 (0.10)	0.06 (0.10)	0.06 (0.07)
Memory control	0.09 (0.17)	−0.02 (0.11)	0.04 (0.08)	0.04 (0.06)

critical question was now worded to tap participants' memory for the original placement of the object: "Then Yoko comes back. Where did she put the ice cream before she went out to play?" Again, the correct response to this question was the first location. Although the false belief and memory control trials are conceptually identical except for the critical question, they differ in a fundamental way: the false belief trials require that participants remember where the protagonist originally placed the object and also to set aside their own privileged knowledge of L2 to respond from the perspective of the naïve protagonist; the memory control trials require only that participants remember where the protagonist originally placed the object.

Participants in the Sandbox task have a tendency to report a location closer to L2 rather than the correct answer, L1. We refer to this systematic response bias as L2 responses. Importantly, in the false belief condition, errors can result from (1) a failure to ignore one's own privileged information, or (2) a failure to remember L1. Based on the false belief condition alone, it is, therefore, impossible to say whether only one or both of these errors occur. To clarify the degree of failure to remember L1, we can examine the magnitude of bias in the memory control condition. Errors made in this condition are independent of false beliefs. Consequently, more errors in the false belief condition compared to the memory control condition indicate a failure in false belief reasoning and egocentric responding, that is, a failure to ignore one's own privileged information.

For both false belief and memory control trials, we varied whether L2 moved to the right or the left of L1. This helped to prevent any inherent negative or positive bias for either the false belief or memory control trials. We also varied whether the distance between L1 and L2 was 35 or 70 mm. We compared bias on the false belief questions to bias on the memory control questions for both the short and long trials. The correct response to every false belief and memory control scenario was L1, while the correct response to the one true belief scenario was L2. The true

belief trial was included strictly to prevent participants from developing a response strategy; therefore, we did not analyze these responses. We expected that participants would show a greater shift towards L2 on the false belief trials than the memory control trials, indicating egocentric perspective taking, resulting from a failure to set aside the privileged knowledge of the second location.

Results

We measured location accuracy as the distance in mm between L1 and the participant's answer. Any response that moved towards L2 we considered a positive bias, while movement in the opposite direction of L2 we considered a negative bias. We divided each bias score by the total length of the Paper and Pencil Sandbox to create a proportion score. Means and standard deviations for long and short trials are included in Table 1. Performance on false belief long trials correlated with performance on false belief short trials $r = 0.30$, $p = 0.007$. Performance on the memory control long trials did not significantly correlate with performance on memory control short trials $r = 0.12$, $p = 0.29$. We compared average bias on the two false belief long trials (70 mm) to the average bias on the two memory control long trials (70 mm). As expected, false belief reasoning bias was greater than memory control bias $t(80) = 3.03$, $p = 0.003$, $d = 0.43$. Next, we compared the average bias on the two false belief short trials (35 mm) to the average bias on the two memory control short trials (35 mm). Again, false belief reasoning bias was greater than memory control bias $t(80) = 5.05$, $p < 0.001$, $d = 0.72$.

These results indicate that with a sufficiently sensitive task, adults demonstrate errors in false belief reasoning similar to those seen in young children on the classic change-of-location task (Birch & Bloom 2007; Maehara & Umeda, 2013; Mitchell, Robinson, Isaacs, & Nye, 1996; although see Ryskin & Brown-Schmidt, 2014). We ran additional analyses to ensure that the results were not due to a few participants showing extreme scores on specific trials. We evaluated each individual's overall bias score by subtracting overall memory control from overall false belief reasoning bias. Sixty-three participants (77 % of the sample) showed more false belief than memory control bias on this task as indicated by positive difference scores. Chi-square analysis comparing those who did and did not show bias was significant $\chi^2(1, N = 81) = 25.00$, $p < 0.001$. As a measure of individual differences on our task, we examined the number of participants who performed without error. We arbitrarily defined "errorless" performance for the long trials and short trials separately as follows: average bias falling between -0.013 and $+0.013$ (proportion) for the long trials (2.7 % of the total distance

of the sandbox), and average bias falling between -0.007 and $+0.007$ (proportion) for the short trials (1.35 % of the total distance of the sandbox). Five participants (0.06 proportion) exhibited errorless performance on the false belief long trials; 19 participants (0.23 proportion) exhibited errorless performance on the memory control long trials; 13 participants (0.16 proportion) exhibited errorless performance on the false belief short trials; 0 participants exhibited errorless performance on the memory control short trials.

Discussion

Most previous attempts to measure adult first-order false belief reasoning bias have failed because tasks designed for children were insensitive to adult performance. The results from Experiment 1 resemble those of Keysar et al. (2003) as well as Surtees and Apperly (2012) who both argue that adults demonstrate perspective-taking errors. Egocentric perspective taking may be an inherently human characteristic that is not only present in children, but also in adults. The Paper and Pencil Sandbox task adds to a growing and necessary arsenal of tasks for measuring ToM in different ages (Devine & Hughes, 2013; Lagattuta, Sayfan, & Harvey, 2013; Surtees & Apperly, 2012). Our findings suggest that previous conceptions of ToM suddenly maturing at age 4 may reflect the categorical nature of standard ToM tasks. The Paper and Pencil Sandbox task may be more appropriate than a categorical belief reasoning task for measuring ToM performance in relative degrees rather than absolute terms. Also, researchers could potentially use the Paper and Pencil Sandbox task to measure first-order false belief reasoning bias in other language-competent age groups (e.g., preschoolers through older adults) without having to adjust the materials or procedures (cf., Bernstein, Erdfelder, Meltzoff, Peria, & Loftus, 2011).

Examining trials of errorless performance revealed that 19 individuals showed no errors on the memory control long trials. However, there were no observed errorless memory control short trials. Rather, individuals had a tendency to show a negative bias on the short trials. This could suggest that the difference between false belief and memory control short trials resulted from individuals' failure to remember the original location rather than a failure in perspective taking. In Experiment 2, we sought to replicate and extend the effects that we observed in Experiment 1. Specifically, in Experiment 2, we tried to validate the Paper and Pencil Sandbox task by comparing it to performance on the Real Object Sandbox task. Also, we examined whether participants would exhibit a pattern of errorless performance similar to what we observed in Experiment 1.

Experiment 2

To demonstrate that the Paper and Pencil Sandbox task and the Real Object Sandbox task tap a similar construct, we tested correlations between the two tasks. In particular, correlations between false belief and memory control performance on the Paper and Pencil and the Real Object version would add to the validity of the Paper and Pencil Sandbox task. Additionally, we sought to replicate procedures from Experiment 1 using only short trials. More positive bias on false belief compared to memory control trials would indicate that bias is driven by a failure in perspective taking rather than memory.

Method

Participants

Eighty-five undergraduates from a Canadian university (*M* age 21.76 years; range 17–44 years; 83 % female) received course credit through the psychology research pool.

Materials

Paper and pencil sandbox task We developed a new Paper and Pencil Sandbox task with novel scenarios to allow participants to complete both the Paper and Pencil and the Real Object version of the Sandbox task. This prevented the participants from hearing identical scenarios in the two versions of the Sandbox task. Additionally, because the Real Object Sandbox task holds the distance between locations constant, the distance between L1 and L2 on the Paper and Pencil Sandbox task in Experiment 2 was constant on all trials. We chose to hold the distance between L1 and L2 at 35 mm because this distance yielded a larger effect size than the 70 mm distance in Experiment 1, and we wanted to examine if the pattern of errorless performance on memory control trials that we observed in Experiment 1 would replicate. Finally, to control for possible order effects, we counterbalanced whether the participants started with a false belief or memory control trial. We blocked the false belief and the memory control trials; therefore, participants completed four memory control trials or four false belief trials in succession before completing a true belief trial and then four false belief or four memory control trials, respectively. We also counterbalanced whether the participants started with the Real Object Sandbox task or the Paper and Pencil Sandbox task.

Real object sandbox task The Real Object Sandbox task involved a 5-foot (150 cm) long Styrofoam-filled box (Sommerville et al., 2013). The procedures were similar to those for the Paper and Pencil Sandbox task described in Experiment 1 with the following exceptions: the participants stood facing the researcher who used real objects (approximately 50–75 mm in length) to depict each scenario; the objects were buried in the Styrofoam-filled sandbox which was placed between the researcher and the participant; and the distance between L1 and L2 was always 35 cm. After hearing the first part of the scenario, the participant turned around to complete the 20-s filler task. When the participant turned back around the researcher asked the critical question, to which the participant responded by pointing into the sandbox. The researcher then used a measuring tape that was attached to the opposite side of the sandbox (concealed from the participant) to note the response on each trial. To control for possible order effects, we counterbalanced whether the participants started with a false belief or memory control trial. If participants started with memory control (or false belief), they did so for both the Paper and Pencil, and the Real Object Sandbox task.

Design

We used a 2 (belief: false belief, memory control) × 2 (task: Paper and Pencil, Real Object) × 2 (task order: Real Object first, Paper and Pencil first) × 2 (belief order: false belief first, memory control first) mixed design. Belief and task were within-subjects variables and task order and belief order were between-subjects variables.

Procedure

Participants provided full informed consent before completing this study. Procedures were similar to those described in Experiment 1. Each participant completed nine trials of either the Paper and Pencil Sandbox task or the Real Object Sandbox task. They then performed an unrelated filler task before completing nine different trials on the corresponding task. We expected a main effect of belief as evidenced by greater errors on false belief trials than memory control trials. We did not expect to observe an effect of task or a significant belief by task interaction.

Results

As in Experiment 1, we measured location accuracy as the distance between L1 and the participant's answer. Any response that moved towards L2 we considered a positive bias, while movement in the opposite direction of L2 we considered a negative bias. We divided each bias score by the total length of the Sandbox (Paper and Pencil or Real Object) to create a proportion score. This allowed us to compare bias on the two tasks. We calculated overall false belief reasoning bias and overall memory control bias by

averaging the four false belief trials on each task and by averaging the four memory control trials on each task, respectively.

There were significant correlations between false belief performance on the Paper and Pencil Sandbox task and the Real Object Sandbox task, $r = 0.43$, $p < 0.001$, as well as memory control performance between the Paper and Pencil Sandbox task and the Real Object Sandbox task, $r = 0.39$, $p < 0.001$. A 2 (belief) × 2 (task) × 2 (task order) × 2 (belief order) Analysis of Variance was conducted to examine false belief performance across the two tasks. As predicted, there was a main effect of belief $F (1, 80) = 5.19$, $p = 0.03$, $\eta_p^2 = 0.06$. Participants showed more bias on false belief trials ($M = 0.06$, SD $= 0.07$) than memory control trials ($M = 0.04$, SD $= 0.06$). The main effect of task was not significant $F (1, 80) = 0.34$, $p = 0.57$, nor was the belief by task interaction $F (1, 80) = 0.11$, $p = 0.74$. There was a significant task-by-task-order effect $F (1, 80) = 17.09$, $p < 0.001$. Participants showed less bias overall (false belief and memory control collapsed) on whichever Sandbox task they performed second. This was more pronounced on the Paper and Pencil Sandbox task ($M = 0.03$, SD $= 0.11$ when performed second; $M = 0.07$, SD $= 0.10$ when performed first) than on the Real Object Sandbox task ($M = 0.04$, SD $= 0.08$ when performed second; $M = 0.06$, SD $= 0.08$ when performed first). There were no other significant interactions. Means and standard deviations for the Paper and Pencil Sandbox and the Real Object Sandbox are included in Table 1.

As a measure of individual differences on our Paper and Pencil Sandbox task in Experiment 2, we examined the number of participants who performed without error. As in Experiment 1, we arbitrarily defined "errorless" performance as average bias (this time across all four trials) falling between -0.007 and $+0.007$ (proportion). Three (0.04 proportion) participants exhibited errorless performance on the false belief trials; nine (0.11 proportion) participants exhibited errorless performance on the memory control trials. As a measure of individual differences on our Real Object Sandbox task, we examined the number of participants who performed without error. Again, we arbitrarily defined "errorless" performance as average bias (across all four trials) falling between -0.007 and $+0.007$ (proportion). Eleven (0.13 proportion) participants exhibited errorless performance on the false belief trials; 12 (0.14 proportion) participants exhibited errorless performance on the memory control trials. One participant exhibited errorless performance on both the Paper and Pencil Sandbox and the Real Object Sandbox false belief trials (0.01 proportion). Three participants exhibited errorless performance on both the Paper and Pencil

Sandbox and the Real Object Sandbox memory control trials (0.04 proportion).

Discussion

Significant correlations emerged between the Paper and Pencil Sandbox task and the Real Object Sandbox task. This suggests that the two Sandbox tasks are tapping a similar construct. Research comparing performance on different ToM tasks is informative and contributes to the literature. Correlations between ToM tasks are not always found and some research indicates that different tasks tap different skills (Brent, Rios, Happé, & Charman, 2004; Henry, Phillips, Ruffman, & Bailey, 2013). For example, some tasks, such as the Reading the Mind in the Eyes task (Baron-Cohen et al., 2001), appear to tap more emotional recognition or affective processing, while other tasks, such as the Sandbox tasks tap more cognitive perspective taking (Henry et al., 2013). Additionally, results from a meta-analysis suggest that there is a relationship between performance on intelligence tasks and the Reading the Mind in the Eyes task (Baker, Peterson, Pulos, & Kirkland, 2014). This is but one example of research indicating that different ToM tasks could also be assessing various general cognitive abilities. Future research could consider running both versions of the Sandbox task with other advanced ToM tasks, such as the Strange Stories task (Happé, 1994), or the Reading the Mind in the Eyes task (Baron-Cohen et al., 2001). Investigation of how different ToM tasks relate to one another could further our understanding of ToM development.

The results from Experiment 2 show that on both the Paper and Pencil Sandbox task, as well as the Real Object Sandbox task, participants have a tendency to shift more towards the outcome location (L2) when they are taking the perspective of the protagonist (false belief) compared to when they are simply remembering object location (memory control). There was one significant interaction, a task-by-task-order effect, suggesting that individuals showed less bias overall (higher accuracy) on whichever Sandbox task they performed second. The effect was more pronounced when participants completed the Paper and Pencil Sandbox second, than when they completed the Real Object Sandbox task second. This was most likely due to practice effects being more pronounced on the Paper and Pencil Sandbox task than on the Real Object Sandbox task. Additionally, it could be that performance on the Real Object Sandbox task is more stable than the version of the Paper and Pencil Sandbox task used in Experiment 2. Future research could seek to confirm this speculation.

General discussion

The Paper and Pencil Sandbox task is a convenient version of the Real Object Sandbox task. Both tasks provide continuous measures of false belief performance sensitive enough to detect perspective-taking bias in adults. In Experiment 1, participants showed more bias on false belief trials, which required perspective taking, than on memory control trials, which required memory for the original object location. Therefore, participants were more likely to shift their response to L2 (the outcome location) when they had to take the perspective of the protagonist than when they simply had to remember the object's original location. In Experiment 2, we sought to validate the Paper and Pencil Sandbox task by having participants complete it along with the Real Object Sandbox task. The Paper and Pencil Sandbox task correlated with the Real Object Sandbox task, suggesting that the two tasks tap a similar construct.

There were some differences in performance on the Paper and Pencil Sandbox task between the two experiments. Specifically, the proportion of bias was greater in Experiment 1, where we tested the Paper and Pencil Sandbox task on its own, compared to Experiment 2, where we tested the Paper and Pencil Sandbox task along with the Real Object Sandbox task. We made several changes to the Paper and Pencil Sandbox task in Experiment 2 to allow it to be run with the Real Object Sandbox task: We created nine new scenarios and omitted the long trials from the Paper and Pencil Sandbox task. Also, in Experiment 2, participants completed 18 trials between the two Sandbox tasks rather than the nine trials they completed in Experiment 1. It could be that the task used in Experiment 1 is a more sensitive measure of false belief performance. For example, in Experiment 1 the Sandbox task consisted of both long and short trials. It could be that participants are less able to develop performance strategies when two different trial lengths are used. We also blocked the false belief and memory control trials in Experiment 2 rather than alternating them as in Experiment 1. Additionally, we did not observe the same pattern of errorless trials in both experiments. Recall that in Experiment 1, there were no errorless memory control short trials and the consideration was that the results were being driven by negative bias on the memory control trials instead of a positive bias on the false belief trials. In Experiment 2, nine participants performed without error on the memory control trials and the overall mean bias on memory control trials was positive. In sum, the observed differences in performance on the Paper and Pencil Sandbox task between the two experiments are most likely due to changes in procedures, including changes to the Paper and Pencil Sandbox task itself.

The current findings add to a line of research that highlights false belief reasoning bias in adults (see Royzman, Cassidy, & Baron, 2003). While both social and developmental psychologists have studied this topic extensively, the limited perspective-taking skills of normal adults seem to be inconsistent with many studies from the field of developmental psychology (Piaget, 1966; Tversky & Kahneman, 1974). In developmental psychology, it is often assumed that adults have a "full-blown" ToM after passing the first-order false belief task, suggesting that this capacity is in place much like our sensory information processes (Berk, 2012). Closer collaboration among social, cognitive, and developmental psychologists may add to our understanding of ToM in both children and adults. We believe that the Paper and Pencil Sandbox task could be used as a valuable tool to help bridge these subfields.

The results from both experiments resemble those of Begeer et al. (2012) who demonstrated that typically developing adolescents showed more bias on a single false belief trial than on a single no-false belief trial in a paper and pencil version of the Real Object Sandbox task. The no-false belief trial used by Begeer et al. (2012) depicted the character placing a new object in a second location, rather than changing the location of the original object. In the current study (Experiment 1), the effect size denoting the difference between the false belief trials and memory control trials ($d = 0.43$ for long trials and $d = 0.72$ for short trials) was greater than the effect size between the single false belief trial and the single no-false belief trial reported by Begeer and colleagues ($d = 0.35$). In Experiment 2 of the current study, however, the effect size ($d = 0.23$) was smaller than that observed by Begeer et al. ($d = 0.35$). Additionally, Begeer and colleagues showed differences in performance between typically developing adolescents and adolescents with high functioning autism. Identification of perspective-taking deficits is an important element of clinical assessment, particularly when evaluating and treating developmental disorders. Future research could extend the findings of Begeer et al. (2012) and employ the Sandbox tasks to compare performance of individuals with and without developmental disorders.

The two versions of the Sandbox task, the Paper and Pencil version and the Real Object version could conveniently be administered to various age groups from preschoolers to older adults, without having to adjust the procedures. Either task could be used to explore developmental differences in ToM performance. Future research could investigate whether older and younger adults show performance differences on the Paper and Pencil Sandbox task, as demonstrated with the Real Object Sandbox task (Bernstein, Thornton, & Sommerville, 2011).

In sum, previous change-of-location tasks used only two locations and resulted in ceiling performance in most

typically developing individuals by the age of 6, making these tasks unsuitable for detecting false belief errors in older children and adults (Birch & Bloom, 2007). The current study has demonstrated a convenient Paper and Pencil version of the Sandbox task, successfully measuring adult false belief errors (bias) in situations that require individuals to take another person's perspective. The Paper and Pencil Sandbox task correlated with the Real Object Sandbox task, further establishing its validity as a measure of false belief reasoning. Researchers could use the Paper and Pencil Sandbox task to assess false belief reasoning, and could do so easily and inexpensively. The Paper and Pencil Sandbox task could, thus be a valuable contribution to research on the development of perspective taking.

References

Baker, C. A., Peterson, E., Pulos, S., & Kirkland, R. A. (2014). Eyes and IQ: a meta-analysis of the relationship between intelligence and "Reading the Mind in the Eyes". *Intelligence, 44*, 78–92.

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

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "Reading the Mind in the Eyes" test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry, 42*, 241–251.

Begeer, S., Bernstein, D. M., van Wijhe, J., Scheeren, A. M., & Koot, H. (2012). A continuous false belief task reveals egocentric biases in children and adolescents with Autism Spectrum Disorders. *Autism, 16*, 357–366.

Berk, L. E. (2012). *Child development* (9th ed.). Boston: Allyn and Bacon.

Bernstein, D. M., Erdfelder, E., Meltzoff, A. N., Peria, W., & Loftus, G. R. (2011). Hindsight bias from 3 to 95 years of age. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 37*, 378–391.

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*, 481–502.

Birch, S. A. J., & Bloom, P. (2007). The curse of knowledge in reasoning about false beliefs. *Psychological Science, 18*, 382–386.

Brent, E., Rios, P., Happé, F., & Charman, T. (2004). Performance of children with autism spectrum disorder on advanced theory of mind tasks. *Autism, 8*, 283–299.

Carpendale, J. I., & Chandler, M. J. (1996). On the distinction between false belief understanding and subscribing to an interpretive theory of mind. *Child Development, 67*, 1686–1706.

Devine, R. T., & Hughes, C. (2013). Silent films and strange stories: theory of mind, gender, and social experiences in middle childhood. *Child Development, 84*, 989–1003.

Happé, F. G. (1994). An advanced test of theory of mind: understanding the story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *Journal of Autism and Developmental Disorders, 24*, 129–154.

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*, 826–839.

Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition, 89*, 25–41.

Lagattuta, K., Sayfan, L., & Harvey, C. (2013). Beliefs about thought probability: evidence for persistent errors in mindreading and links to executive control. *Child Development, 85*, 659–674.

Maehara, Y., & Umeda, S. (2013). Reasoning bias in the recall of one's own beliefs in a smarties task for adults. *Japanese Psychological Research, 55*, 292–301.

Miller, S. A. (2009). Children's understanding of second-order mental states. *Psychological Bulletin, 135*, 749–773.

Mitchell, P., Robinson, E. J., Isaacs, J. E., & Nye, R. M. (1996). Contamination in reasoning about false belief: an instance of realist bias in adults but not children. *Cognition, 59*, 1–21.

Nickerson, R. S. (1999). How we know and sometimes misjudge—what others know: imputing one's own knowledge to others. *Psychological Bulletin, 125*, 737–759.

Perner, J., & Roessler, J. (2012). From infants' to children's appreciation of belief. *Trends in Cognitive Sciences, 16*, 519–525.

Perner, J., & Wimmer, H. (1985). "John thinks that Mary thinks that…" Attribution of second-order beliefs by 5- to 10- year-old children. *Journal of Experimental Chld Psychology, 39*, 437–471.

Piaget, J. (1966). *Judgment and reasoning in the child*. Totowa: Littlefield, Adams & Company.

Rakoczy, H., Harder-Kasten, A., & Sturm, L. (2012). The decline of theory of mind in old age is (partly) mediated by developmental changes in domain-general abilities. *British Journal of Psychology, 103*, 58–72.

Royzman, E. B., Cassidy, K. W., & Baron, J. (2003). "I know, you know" Epistemic egocentrism in children and adults. *Review of General Psychology, 7*, 38–65.

Rubio-Fernández, P., & Geurts, B. (2012). How to pass the false-belief task before your 4th birthday. *Psychological Science, 24*, 27–33.

Ryskin, R. A., & Brown-Schmidt, S. (2014). Do adults show a curse of knowledge in false-belief reasoning? A robust estimate of the true effect size. *PLoS ONE, 9*, e92406. doi:10.1371/journal.pone.0092406.

Scheeren, A. M., de Rosnay, M., Koot, H. M., & Begeer, S. (2013). Rethinking theory of mind in children and adolescents with autism. *Journal of Child Psychology and Psychiatry, 54*, 628–635.

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.

Surtees, A. D. R., & Apperly, I. A. (2012). Egocentrism and automatic perspective taking in children and adults. *Child Development, 83*, 452–460.

Tager-Flusberg, H., & Sullivan, K. (1994). A second look at second-order belief attribution in autism. *Journal of Autism and Developmental Disorders, 24*, 577–586.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science, 185*, 1124–1131.

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

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.

Yirmiya, N., Erel, O., Shaked, M., & Solomonica-Levi, D. (1998). Meta-analyses comparing theory of mind abilities of individuals with autism, individuals with mental retardation, and normally developing individuals. *Psychological Bulletin, 124*, 283–307.