Implicit false belief across the lifespan: Non-replication of an anticipatory looking task

Article *in* Cognitive Development · September 2017

DOI: 10.1016/j.cogdev.2017.08.006

Alexa Ruel
Concordia University Montreal

Diane Poulin-Dubois
Concordia University Montreal

Cognitive Development xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Cognitive Development

journal homepage: www.elsevier.com/locate/cogdev

Implicit false belief across the lifespan: Non-replication of an anticipatory looking task

Kimberly Burnside*, Alexa Ruel, Naomi Azar, Diane Poulin-Dubois

Concordia University, 7141 Sherbrooke Street West, Montréal, Québec, H4B 1R6, Canada

A R T I C L E I N F O

Keywords:
Implicit false belief
Theory of mind
Anticipatory looking
Non-replication

A B S T R A C T

Implicit false belief is often measured through anticipatory looking. While past research stipulates that children and adults have a fully developed implicit false belief understanding, there is a lack of consensus in recent findings. The goal of this study was to examine how adults and children perform on an anticipatory looking task to further our understanding of the variability in results across studies. The implicit false belief task featured a 3-second anticipatory looking period, during which we measured participants' looking behavior (first look and total looking time at the correct location). We failed to replicate previous findings, with neither group demonstrating an implicit understanding of false belief. However, performance varied depending on the measure examined, thus highlighting the importance of analyzing several variables when assessing false belief with anticipatory looking.

1. Introduction

Theory of mind is a foundational cognitive ability that makes possible the attribution of mental states to oneself and to others. This includes the aptitude to understand another person's beliefs, to recognize that these beliefs may be different from one's own, and to predict how this person will act based on his/her beliefs (Wimmer & Perner, 1983). The capacity to attribute a false belief to another individual (i.e., understanding that an individual holds a belief that is incongruent with reality) is an essential component of a theory of mind. It was initially proposed that this understanding develops around 4 years of age (Perner, Leekam, & Wimmer, 1987; Wellman, Cross, & Watson, 2001). A wealth of evidence has since confirmed that children before the age of 4 years do not succeed on tasks examining false belief. Over two decades ago, a milestone study by Clements and Perner (1994) provided empirical evidence showing that an implicit understanding might develop earlier, at approximately 2 years and 11 months. To measure false belief implicitly, the authors used a variation of the standard Sally-Anne false belief task (Wimmer & Perner, 1983) and measured the direction of the children's first look. They observed that children understand false belief, when measured implicitly, at a younger age than was traditionally believed.

The innovative method introduced by Clements and Perner (1994) was further developed and evolved into what is now known as the anticipatory looking paradigm. In contrast to the original task, instead of asking the participants where the protagonist will look for the object, these paradigms use an anticipatory looking period, during which the participants' looking behavior is analyzed to determine if they correctly expect/anticipate the protagonist's actions (Southgate, Senju, & Csibra, 2007). The question asked in the original paradigm (i.e., "where will [the protagonist] look for the object") may prompt the children to look (or point) where the

* Corresponding author at: Centre for Research in Human Development, Department of Psychology, Concordia University, 7141 Sherbrooke Street West, Montréal, Québec, H4B 1R6, Canada.

E-mail addresses: kimberly.burnside1@gmail.com (K. Burnside), alexa.ruel@mail.concordia.ca (A. Ruel), naomi.azar@hotmail.com (N. Azar), diane.poulindubois@concordia.ca (D. Poulin-Dubois).

http://dx.doi.org/10.1016/j.cogdev.2017.08.006
Received 13 February 2017; Received in revised form 9 August 2017; Accepted 10 August 2017
0885-2014/

object is actually located before fully processing the question. Southgate et al. (2007) created a video based on an anticipatory looking paradigm and tested implicit false belief understanding in 25-month-old toddlers ($N = 20$). In this video, a ball is placed in one of two boxes located in front of two doors. The protagonist is then distracted while a puppet retrieves the ball and leaves the scene with it. The two doors (areas of interest) light up and the scene freezes for an anticipatory looking period of 1750 ms. Following this period, the protagonist opens the door corresponding to the box in which she left the ball in order to retrieve it. The anticipatory looking period allowed for the investigation of the children's implicit understanding of false belief by examining the direction of the first look as well as the amount of time spent looking at both doors. As expected, a statistically significant proportion of infants (85%) correctly anticipated the protagonist's behavior, directing their first look towards the door above the box where the protagonist left the ball. Additionally, their total looking time to the correct door was almost twice as long as their looking time to the incorrect door during the anticipatory looking period; this difference was significantly above chance. The authors concluded that anticipatory looking tasks convincingly demonstrate that 25-month-old toddlers attribute false belief to another agent by correctly anticipating his/her future behavior.

Another study by Thoermer, Sodian, Vuori, Perst, and Kristen (2012) examined the same question by designing a similar task (the "autobox" task) with a group of forty-four 18-month-old infants. This anticipatory looking task included a self-moving car first shown moving from one garage to another, followed by a 3-second anticipatory looking period. The protagonist then reached for the car by opening the door above the garage corresponding to the car's location. During the test trial, once the car reached the garage, the protagonist was distracted as the car returned towards its initial location and then exited the scene. This was followed by a 3-second anticipatory period after which the protagonist opened one of the two doors in order to reach for the car. A ratio was calculated by dividing the time participants spent looking at the correct door by the sum of looking times at both the correct and incorrect doors. The resulting ratio could range from 0 to 1, where scores above 0.50 represented a preference for the correct door, and scores below 0.50, a preference for the incorrect door. In this study, 55% of 18-month-old infants successfully spent more time looking at the door corresponding to the protagonist's false belief, which reflects an understanding of false belief when using an implicit measure. Although not included in the original paper, when calculated, this success rate was not above chance, which is 50% ($p = 0.120$). In a recent study, the same implicit false belief task was administered to forty-four 18-month-olds who looked towards the correct door during the anticipatory looking period on average 57% of the time (Sodian et al., 2016). Similar to Thoermer et al. (2012), this percentage did not differ from chance ($p = 0.180$).

Schuwerk, Jarvers, Vuori, and Sodian (2016) attempted to replicate the findings reported with this task (Thoermer et al., 2012) with 21 children aged 7–8 years. They sought to compare the performance of children diagnosed with Autism Spectrum Disorder (ASD) to that of neurotypical children. Although it is well documented that children with ASD lack an explicit understanding of false belief (Baron-Cohen, Leslie, & Frith, 1985), the objective of the study was to assess whether children with ASD would also fail false belief when measured implicitly. The authors calculated a differential looking score (DLS) by subtracting the time spent looking at the incorrect door from the time spent looking at the correct door and dividing this difference by the time spent looking at both doors. Thus, the resulting DLS could range from -1 to 1, with negative scores reflecting a preference for the incorrect door and positive scores indicating a preference for the correct door. There was a sharp contrast in the performance of the neurotypical children when compared to the children with ASD. The neurotypical children showed a stronger bias toward the correct door (DLS $= 0.21$) than the children with ASD, who looked longer at the incorrect door (DLS $= -0.23$). Although the neurotypical children showed a stronger looking bias toward the correct door, this DLS was not statistically significantly above chance ($t(20) = 1.63$, $p = 0.118$; Schuwerk, personal communication, 2016). The DLS of the neurotypical children was only significantly above chance when both test trials were merged. Therefore, neurotypical children's performance on this task was weaker than that observed in the studies conducted by Southgate et al. (2007) and Senju, Southgate, White, and Frith (2009).

Although implicit false belief is traditionally assessed in infants and clinical populations (i.e., ASD), there are also some studies conducted on adult samples. Implicit false belief is believed to be a form of false belief tracking and is therefore believed to be stable across the lifespan. Thus, children and adults are expected to perform well on implicit false belief tasks because they should have a fully developed efficient system of false belief tracking (Apperly & Butterfill, 2009). Unexpectedly, the performance of neurotypical adults on anticipatory looking paradigms also varies considerably across studies. Some studies, using the task by Southgate et al. (2007), or variations of this task, reported that at least 90% of adult samples first looked towards the correct area of interest, demonstrating an implicit understanding of false belief (Low & Watts, 2013; Wang, Hadi, & Low, 2015). Similarly, Senju et al. (2009) and Senju (2012) reported that neurotypical adults' first look was directed towards the correct area of interest. However, Wang and Leslie (2016) also assessed adults on the same task and found that only 64% of the adults' first look was directed towards the correct location. Other studies with comparable sample sizes also failed to report high rates of false belief understanding in adults. For example, Schuwerk, Vuori, and Sodian (2015) used the same anticipatory looking task as Thoermer et al. (2012) in an adult sample ($N = 19$). The authors reported that 74% of adults first looked towards the correct location on the first test trial, a result just trending toward statistical significance. However, this decreased to 58% on the second trial. The lower passing rate on the second test trial could be due to a decrease in motivation given that this is the second time that they see the video (in addition to the two familiarization trials). Taken together, these results indicate that absolute rates of success in adult performance on this implicit false belief task are, like in children, inconsistent.

One important issue to consider when comparing results obtained from anticipatory looking tasks is the differences in task demands. Indeed, anticipatory looking paradigms can be divided into tasks of high vs. low demand (Wang & Leslie, 2016). In a high demand anticipatory looking paradigm, the object remains in the scene after being moved to a different location, as in the traditional Sally-Anne task (Baron-Cohen et al., 1985; Grosse Wiesmann, Friederici, Singer, & Steinbeis, 2016; Wang & Leslie, 2016). This task has more cognitive demands because the knowledge about the current location of the object must be inhibited as it is in competition with the alternative belief of the protagonist. Conversely, in a low demand task (e.g., Southgate et al., 2007; Thoermer et al., 2012), the object is removed from the scene. Thus, the need for the viewer to inhibit looking towards the actual location of the object (i.e., the pull of reality) is removed, which in turn facilitates the attribution of an alternate belief to the protagonist. Although this explanation seems plausible when attempting to explain the observed variability in results, some studies fail to support this hypothesis. For instance, given that the autobox task used by Schuwerk et al. (2015) was a low demand task (i.e., the car left the scene), adults' performance should be superior to that of a high demand version of the task. However, the adults' performance was better on the high demand anticipatory looking task in the study conducted by Low and Watts (2013) than performance on the low demand task of Schuwerk et al. (2015).

Only recently have high and low demand implicit false belief tasks been directly compared by slightly modifying the Southgate et al. (2007) paradigm (Wang & Leslie, 2016). Adults completed a high ($N = 42$) and low demand ($N = 44$) version of the task, as did toddlers between 2 and 4 years of age (high: $N = 29$, low: $N = 27$). During the low demand false belief task, 64% of the adults and 70% of the toddlers directed their first look towards the false belief window (correct window). Consistent with the authors' hypothesis for the high demand version of the task, first look accuracy was not statistically significant with 48% of adults and 55% of toddlers looking at the correct window first. It is also important to note that the authors did not use the same exclusion criteria as Southgate et al. (2007). Instead of excluding participants who did not pass the second familiarization trial, the authors excluded the participants who did not pay sufficient attention to either window (in terms of total looking time). Taken together, the results of the studies reported above do not show consistency in adults' and children's performance on comparable versions of the anticipatory looking paradigm.

In the current study, we attempted to address the issue of replicability of the anticipatory looking task by examining children's and adults' performance on the same task (i.e., "autobox" task; Schuwerk et al., 2015; Schuwerk et al., 2016; Thoermer et al., 2012). As this task is considered a low demand anticipatory looking false belief task, both groups should demonstrate an implicit understanding of the protagonist's false belief through anticipatory looking. Children from a wide age range were tested in order to assess the robustness of this form of belief tracking in children, regardless of age. We also examined adults' performance on the same anticipatory looking task to confirm that hypothesis across the life span. Given that most of the previous studies that used this task reported small-to-moderate sample sizes (Schuwerk et al., 2015; Schuwerk et al., 2016; Thoermer et al., 2012), another goal of the present study was to use the same low demand paradigm on large samples of children ($N = 63$) and adults ($N = 99$). By doing so, we aimed to provide a strong replication test of a false belief task based on anticipatory looking.

2. Method

2.1. Participants

A total of 77 children participated in this study (41 boys and 36 girls; $M_{\text{age}} = 4.30$ years, range = 2.08–7.50 years). Forty-six participants were tested in English and 31 were tested in French. Fourteen participants were excluded from the analyses because the task was not recorded ($n = 4$), the participant did not calibrate ($n = 4$), the participant did not look at the screen during the test trial ($n = 1$), or the participant was non-compliant ($n = 5$). Thus, the final analyses were conducted on a final sample of 63. Participants were recruited from a University database. The goal was to administer an implicit false belief task to a large sample of children in order to have enough power to detect small-to-moderate effects.

Additionally, a total of 104 adults participated in this study (25 males and 79 females; $M_{\text{age}} = 23.7$ years, range = 18.0–51.8 years). A sample of 101 participants was tested in English and three in French. Participants were recruited from a University participant pool. A total of 59 participants viewed the original implicit false belief video developed by Thoermer et al. (2012) and 45 participants viewed an accelerated version. These two versions were used to assess whether small methodological differences in the task would affect participants' performance, and possibly explain the inconsistency across studies. Five participants were excluded from the analyses because the participant did not calibrate ($n = 2$), or because the participant did not look at the screen during the test trial ($n = 3$). Thus, the final analyses were conducted on a final sample of 99 participants.

2.1.1. Materials and procedure

Adult and children participants were scheduled for one visit to the laboratory. Children's caregivers received $20 as compensation for their child's participation, while adults received 0.5 course credit as compensation or entered a draw to win one of three $25 cash prizes. The participants or the participants' caregivers completed a short demographic questionnaire before the testing session.

The implicit false belief task was administered using a Tobii TX300 eye-tracker on a 23-inch monitor with a 1920 × 1080 pixel resolution. Participants' eye-gaze was analyzed using Tobii Studio (Tobii Technology, Stockholm, Sweden). The average accuracy of this eye-tracker is in the range of 0.4–1° when viewing the screen at a distance of 65 cm. This corresponds to an average error of 12–30 mm on the screen. This eye-tracking system used a time-to-tracking recovery, where the recovery was immediate following eye-blinks and 10–165 ms after lost tracking. The eye-tracking system sampled data at a rate of 120 Hz. Participants were seated 60–70 cm from the screen with a 30° visual angle. Participants' gaze was calibrated with a five-point calibration procedure. Four points were located at each corner of the screen and one point at the center of the screen. The areas of interest were 355 × 325 pixels in width and height.

The implicit false belief task was adapted from a task originally designed for infants (Thoermer et al., 2012) that has since been used to test children and adults. Participants viewed three videos: two familiarization videos and a test video assessing false belief understanding. Prior to each trial, an attractive attention-getter paired with a chime directed children's attention to the screen. Since the original authors reported that children showed low interest in the video relative to infants (Sodian, personal communication, 2014), an accelerated version of the original false belief video was created for this sample. In the original version, the familiarization trials lasted 32 s and the test trial lasted 41 s. In this modified version, the familiarization trials lasted 26 s and the test trial lasted 35 s.

During the familiarization trials, a human agent was seen at the top-center of the screen with two garages on either side of the screen and a door above each garage. The protagonist watched a toy car move from one garage to the other. The two familiarization trials displayed the car moving in each direction: from left to right in one trial and from right to left in the other trial. The order of these familiarization videos was counterbalanced across participants. Following the car's arrival in the garage on the opposite side of the screen, the protagonist disappeared from the scene, hiding behind closed doors. The start of the 3-second anticipatory looking period was signalled by a chime paired with the two doors turning bright red. Following this, the protagonist appeared in the door above the garage containing the toy car and retrieved it. The test trial began as in the familiarization trials, with the car moving from one garage to the other. Half the participants viewed the car moving from the left to the right garage and the other half viewed the car moving from the right to the left garage. However, once the car reached the garage on the opposite side, a phone ring distracted the protagonist, who turned to face the side of the screen. While the protagonist was looking away, the car reversed its direction and exited the scene on the opposite side (i.e., the side where it was initially located). The video continued as in the familiarization trials, with the protagonist disappearing and the doors turning bright red for three seconds. Finally, the protagonist opened the door above the garage where she had last seen the car (i.e., before the distraction).

The two measures used to assess false belief understanding were the direction of the first look to the doors and DLS. In terms of the first look measure, participants passed a trial if their first fixation during the anticipatory looking period was directed towards the door where the protagonist subsequently appears. Looks shorter than 80 ms were excluded from the analyses. Therefore, in the familiarization trials, the direction of the correct first look should be towards the door above the garage where the car is located. In the test trial, the direction of the correct first look was towards the door above the garage in which the car was initially heading before reversing and exiting on the opposite side. Therefore, a passing score was awarded if the participant was able to correctly anticipate the protagonist's action during the test trial, as guided by the understanding of the protagonist's false belief. In line with previous inclusion criteria used with this anticipatory looking task, participants were included if they passed at least one familiarization trial (Schuwerk et al., 2015; Thoermer et al., 2012). Fifty-eight children (92%) passed at least one familiarization trial based on the first look criterion and were therefore included in the final analyses on the first look variable. Seventy-three adults (74%) passed at least one familiarization trial based on the first look criterion and were therefore included in the final analyses. The DLS during the 3-second anticipatory looking period was calculated ([time spent looking towards the correct door − time spent looking towards the incorrect door]/total looking time towards both doors) as described by Senju et al. (2009). The DLS for each participant could range from −1 to 1, where a score of zero indicated that the participant spent equal time looking towards both doors. Therefore, a passing score was awarded if the participant's DLS was above zero, which indicated more time spent looking at the correct door, and thus correct anticipation of the protagonist's action. Fifty-four children (86%) passed at least one familiarization trial based on the DLS criterion and were therefore included in the final DLS analyses. Eighty-six adults (87%) passed at least one familiarization trial based on the DLS criterion and were therefore included in the final DLS analyses.

2.2. Results

Children's chronological age and DLS were normally distributed and the data did not include any outliers. Adults' DLS were normally distributed and did not contain any outliers. Using a z-score cut-off of ± 3.0, there were three outliers based on adults' chronological age. However, given that these adults were not outliers on the implicit false belief measures and that participants' age was not a variable included in the analyses, these participants were included in the final analyses. Normality was not assessed for first look, as it is a dichotomous variable.

When the two groups (children and adults) were combined, 48% of the participants' first look was directed towards the correct door in the test trial. The merged group spent equal time looking at both doors during the test trial ($M_{DLS} = 0.04$, $SD = 0.73$). Additionally, based on the DLS criterion, 50% of the participants passed the implicit false belief task. Group comparisons were conducted to determine if the children and the adults performed similarly on the implicit false belief task. When using the first look variable, the two groups' performances did not statistically differ ($\chi^2(1, N = 129) = 3.05$, $p = 0.08$, $V = 0.15$). However, when comparing the two groups' DLS, there was a statistically significant difference ($t(135) = -2.84$, $p = 0.01$, $d = -0.51$). Following this, we ran separate analyses for each group to further explore children's and adults' performance on the implicit false belief task.

Only 38% of the children's first look was directed towards the correct door at test. A non-parametric binomial test revealed that this proportion tended to be statistically below than what would be expected by chance ($p = 0.09$). Children's DLS was statistically lower than zero ($M = -0.20$, $t(51) = -2.19$, $p = 0.03$, $d = -0.30$), indicating that they spent more time looking at the incorrect door during the anticipatory looking period. Based on the DLS criterion, only 33% of the children passed the implicit false belief task, and this proportion was statistically below chance when using a binomial test ($p = 0.02$). Overall, the children failed the implicit false belief task as measured by both first look and looking time (DLS). In order to assess whether the two variables (first look and DLS) used to measure children's false belief understanding were related, an independent samples *t*-test was conducted. Participant's first look during the anticipatory period was related to their DLS ($r_{pb} = 0.66$, $n = 51$, $p < 0.001$). The children who failed the implicit false belief task based on first look had a mean DLS of −0.52, while the children who passed had a mean DLS of 0.36. This indicates that both measures (first look and DLS) are consistent with one another.

Given that the majority of the children failed the implicit false belief task, post-hoc analyses were conducted to determine where

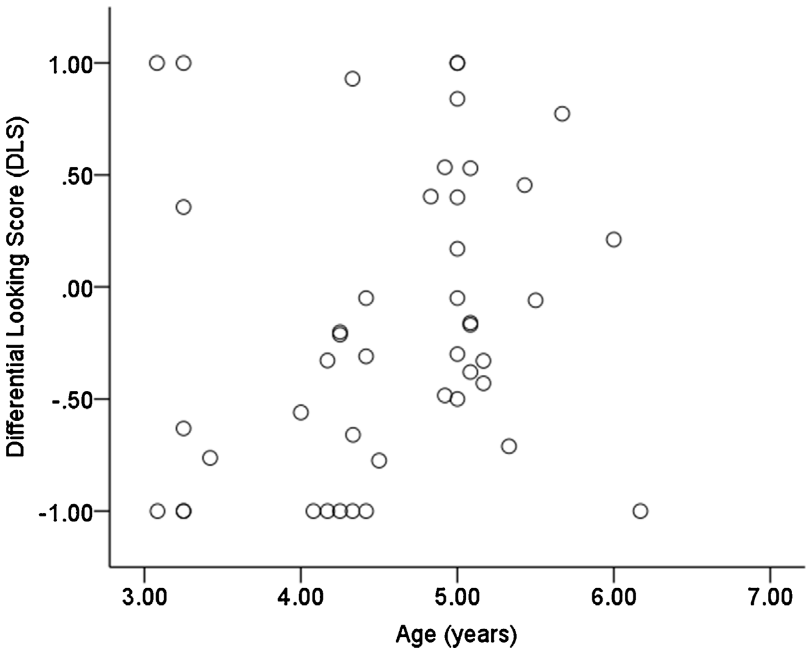

Fig. 1. Children's differential looking score (DLS) as a function of chronological age.

the children were looking during the anticipatory looking period. Specifically, these analyses aimed to determine if they were following a simple rule when watching this video. We hypothesized that children may have followed a recency rule, wherein the direction of their first look would be consistent with the correct location in the last familiarization trial. Of the 36 children who failed, 55% ($n = 20$) incorrectly looked at the door that was correct in the second familiarization trial ($p = 0.22$). Thus, children did not seem to be following a recency rule based on the last familiarization trial. This suggests that the children who failed to demonstrate false belief understanding during the test phase tended to look where the car had disappeared.

Since this task was a low demand task, the children's poor performance cannot be explained by high task demands. Furthermore, as the children included in the analyses passed at least one familiarization trial, their poor performance cannot be explained by a poor understanding of the task. These results do not replicate previous findings on children's implicit false belief tasks based on anticipatory looking (Grosse Wiesmann et al., 2016; Low, 2010; Schuwerk et al., 2015; Southgate et al., 2007; Wang & Leslie, 2016). For example, Schuwerk et al. (2016) found that older children ($M_{age} = 7.2$ years) spent more time looking at the correct door ($M_{DLS} = 0.21$) using the same implicit false belief task, procedure, and stimuli as the one used in the present study. Although our sample includes children from a wider age range than in the original study, no relation was observed between age and performance, $r = 0.21$, $n = 52$, $p = 0.13$, see Fig. 1). Furthermore, when the age range was restricted to 3–5 years of age, the correlation between age and performance was also not significant ($r = 0.21$, $n = 45$, $p = 0.18$). This suggests that children's performance on this implicit false belief task is stable over the preschool period but is not as robust as expected. Finally, a post-hoc power analysis was conducted using G*Power (Faul, Erdfelder, Buchner, & Lang, 2009). The achieved power, with a final sample of 63 children, was 0.76.

Given that two versions of the task were used with the adult sample (i.e., the original version and the accelerated version), analyses were conducted to determine if this methodological difference would lead to different performances. When the first look pass/fail criterion was used, adults' performance across the two versions of the task did not statistically significantly differ ($\chi^2(1, N = 73) = 0.74$, $p = 0.39$, $V = 0.10$). When the DLS pass/fail criterion was used, adults who viewed the accelerated version did not differ from the adults who viewed the original version ($\chi^2(1, N = 86) = 1.70$, $p = 0.19$, $V = 0.14$). Finally, the adults' DLS did not statistically differ across the two versions ($t(84) = 0.42$, $p = 0.68$, $d = -0.09$). This pattern of results indicates that adults who viewed an accelerated version of the video did not differ in their looking behavior compared to adults who viewed the original video. Therefore, both adult samples were collapsed in all further analyses.

A total of 55% of the adults directed their first look towards the correct door during the test trial, a proportion not statistically above chance (binomial test, $p = 0.483$). However, adults' DLS indicated that they spent more time looking at the correct door during the anticipatory looking period ($M = 0.18$, $t(85) = 2.20$, $p = 0.03$, $d = 0.24$). Furthermore, a "pass" based on the DLS criterion was coded if the participant's DLS was above 0. Based on this DLS criterion, 59% of the adults passed the implicit false belief task; this proportion was not statistically above chance (binomial test, $p = 0.11$). An independent samples *t*-test was conducted to compare adults' first look to their DLS. Participants' first look during the anticipatory period was statistically significantly related to their DLS ($r_{pb} = 0.62$, $n = 71$, $p < 0.001$). The adults who failed the implicit false belief task based on first look had a mean DLS of -0.31, while the adults who passed had a mean DLS of 0.62. This indicates that both measures (first look and DLS) are consistent with one another. Given that the adults, as a group, performed differently on these two measures (i.e., performed at chance when first look is used vs. passed when DLS is used), this statistically significant relation sheds light on the discrepancy between first look and DLS when group results are reported. Post-hoc analyses were conducted to determine where the failers were looking during the anticipatory looking period. Of the 33 adults who failed, 15 (45%) incorrectly looked at the door that was the correct one during the second familiarization trial ($p = 0.08$). Therefore, adults, like the children, did not use a recency rule to guide their looking behavior on this task.

It was hypothesized that adults would pass this anticipatory looking task, given that this ability is expected to be stable across the lifespan. This was partially supported by the results: the adults performed at chance on the first look measure and there was a low percentage of passers based on DLS. However, the adults looked longer at the correct door, indicating that their performance reflects some understanding of the protagonist's false belief. It does appear that, unlike children, adults were able to disengage from the location where the car exited and focus on the protagonist's belief. Nevertheless, adults' performance was inferior to that reported in other studies (Schuwerk et al., 2015; Wang & Leslie, 2016). This poor performance cannot be explained by high cognitive demands or methodological changes. Although the present results do not replicate previous findings using the same task (Schuwerk et al., 2015), the percentage of passers based on participants' first look is in line with the percentage reported in the study by Wang and Leslie (2016). The authors used a one-tailed binomial test to assess whether the percent of adults who passed (64%) was above chance whereas we used a two-tailed binomial test, which is more conservative. Therefore, the differences in statistical analyses used might explain the difference in the conclusions across studies.

Additional analyses were conducted to investigate group differences on the three measures of the false belief task: first look pass/fail, DLS pass/fail, and DLS. The two groups tended to differ on their first look during the onset of the anticipatory period ($\chi^2(1, N = 131) = 3.69$, $p = 0.06$, $V = 0.17$); more children tended to fail based on the first look criterion. This finding reiterates that neither group passed the task based on the first look pass/fail criterion. When the DLS pass/fail criterion was used, the two groups were statistically significantly different, wherein more adults passed the task compared to children ($\chi^2(1, N = 107) = 4.59$, $p = 0.03$, $V = 0.21$; see Fig. 2). Finally, when the two groups' DLS were compared, the adults' DLS was statistically significantly higher than the children's DLS ($t(105) = 2.18, p = 0.03, d = -0.56$). These two findings indicate that when the participants' looking time during the entire 3-second anticipatory period is analyzed, adults perform better than children. The post-hoc power analysis revealed that the observed power, with a final sample of 99, was 0.77.

3. Discussion

The main goal of the current study was to investigate children's and adults' performance on the version of the anticipatory looking false belief task (i.e., the "autobox" task) developed by Thoermer et al. (2012). This false belief task is qualified as a low demand task given that the object central to the false belief attribution leaves the scene during the anticipatory looking period. In contrast, a high demand version of this task would require participants to inhibit their tendency to look at the current object location in order to pass the task since the object is still present. To date, it has been difficult to successfully compare children's and adults' performance across studies given the variety of implicit false belief tasks used in the literature.

Given that the testing conditions and the stimuli used were selected to enhance participants' performance, the two age groups tested, namely children and adults, were expected to demonstrate a fully developed implicit grasp of false belief since infants typically pass a similar low demand task. Finally, large samples were tested in order to maximize statistical power. Despite these attempts to maximize participants' performance, neither group passed the task on all three measures analyzed, failing to demonstrate a robust implicit understanding of false belief. The methodology used and the areas of interests defined for the eye-tracking system were carefully selected and made comparable to the methods and settings used by Schuwerk et al. (2016) as an attempt to replicate

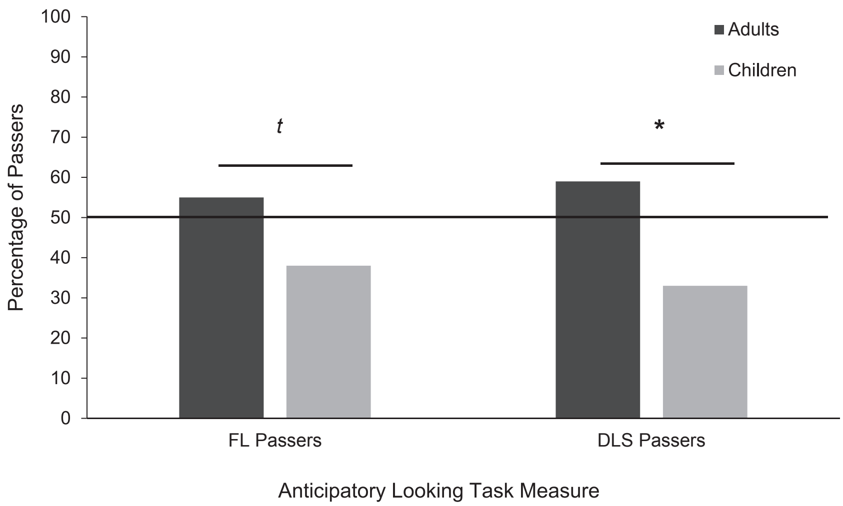

Fig. 2. Percentage of participants who passed the false belief task based on the first look (FL) and DLS criteria for adults and children.
*p < 0.05.

the authors' findings. These researchers used the same task on children and found that they passed the task based on both first look and DLS. Furthermore, Schuwerk et al. (2015) found that adults demonstrate an implicit false belief understanding when the same task was administered.

There are several possible explanations for this failure to replicate. For example, there is variability in the criteria used across studies to assess performance: some authors use one-tailed tests to obtain significance (Wang & Leslie, 2016), while others discuss results in terms of group comparisons rather than testing the passing rate (Schuwerk et al., 2016). That is, using a more lenient statistical test may have inflated the alpha value for the passing rate when compared to a more conservative approach. Importantly, although the task demands in the present version of the anticipatory looking paradigm are low (i.e., the target object is removed), other parameters of the task might contribute to both children's and adults' poor performance. More specifically, the car leaves the scene from one side of the screen, which may bias the participants to look in that direction, that is, at the incorrect door (i.e., expecting the car to return). In the task designed by Southgate et al. (2007), the puppet disappears from the middle of the scene, preventing a bias to look at a particular side of the screen during the anticipatory period. In other words, it is possible that this aspect of the task (i.e., car leaving from a specific side) might make the present task a high demand task.

One of the goals of the present study was to compare the current results obtained with those obtained from other laboratories using the same task. The preschool children in the present study appear to be performing worse than the 7-year-olds in the study conducted by Schuwerk et al. (2016). Similarly, the adults appear to be performing worse than those in the study conducted by Schuwerk et al. (2015). We also introduced small methodological differences (i.e., using both the original version and an accelerated version) in our design to directly test the hypothesis that these differences might explain variable performances across studies. Given that there were no differences in performance across the two video versions, other factors may be responsible for the variability observed across studies.

Of critical importance is the reporting of different measures of implicit false belief when analyzing performance in an anticipatory looking paradigm. One can report the percent of passers based on the direction of the first look, the DLS (i.e., preferential looking to the correct or incorrect location), or even the percent of passers based on DLS. All measures were reported in the present study to illustrate how performance on an anticipatory looking task can be interpreted differently based on the selected measure. That is, reporting participants' first look or DLS can lead to a different passing rate on the same task. For instance, although the children failed the anticipatory looking task according to the three anticipatory looking measures, the adult sample showed a different pattern of results. If solely looking at the adults' average DLS, one could conclude that the adult sample passed the task. However, when looking at the percent of passers based on first look or DLS, the adults perform at chance on this task. A possible explanation as to why DLS seems to better reflect adults' understanding of false belief is that perhaps adults are better able to disengage from the location where the car left the scene (incorrect door) after the initial look at the display.

The adult and child samples performed comparably on this task based on the first look criterion, as both failed on this measure. A possible explanation for this poor performance is that this task fails to robustly assess implicit theory of mind through a measure of first look. Given that the participants' first look arguably assesses their automatic evaluation of the situation, and as both samples fail to demonstrate an understanding of the task using this measure, the task may lack robustness in assessing an automatic understanding of theory of mind. Differences between the stimuli used, particularly the presence of a self-propelled object, in this study and in other low demand paradigms may be driving this effect. In the present task, the apparent self-initiated movement of the car during its exit from the scene on one of two sides may have led the participants to believe that it would return as the protagonist is about to look for it. Further, the current task only includes one agent (i.e., the protagonist), whereas the task used by Southgate et al. (2007) includes two agents (i.e., the protagonist and the puppet). These differences in both versions of a comparable task may have had an impact on the results, particularly on children who have less inhibitory control abilities. Of the participants who failed, the children and adults seem to attempt to make sense of the situation differently. Overall, the methodological differences between the two anticipatory looking paradigms could create different cognitive loads, with the current paradigm by Thoermer et al. (2012) having higher task demands.

As the adults' performance differs across the measures of first look and DLS, one possibility is that these two measures are not reflecting the same understanding of the task. The relatively long duration of the anticipatory looking period (3s) leads to the possibility that the DLS no longer reflects an automatic implicit understanding of theory of mind. Indeed, Southgate et al. (2007) defined their anticipatory looking period by a shorter duration: 1750 ms. Perhaps the DLS is measuring an implicit, although non-automatic, processing of the task. As the adults performed better than the children according to both DLS and DLS pass/fail, it is also possible that the adults were distracted by the car leaving the scene on one side (and thus performed worse on the first look variable), but were still able to disengage and spend more time looking at the correct location (i.e., better performance on the DLS variable) for the rest of the anticipatory period.

The two-systems theory posits that there is an efficient system that is quick, automatic, and does not rely on cognitive ability (i.e., implicit false belief), and a slower, flexible system that requires cognitive input and develops later in childhood (i.e., explicit false belief; Apperly & Butterfill, 2009; Carruthers, 2016; Low, Apperly, Butterfill, & Rakoczy, 2016). Given that explicit false belief understanding was not measured in the current study, it is not possible to comment on the developmental relation between the two systems. At first glance, the present results do not seem to provide support for one of the signatures of System 1: an efficient system that is stable throughout the lifespan. As Carruthers (2016) stated, there is little empirical evidence for this theory and thus, the author encouraged researchers to investigate this theory in a scientific framework. Although not the initial goal, Low and Watts (2013) provide some evidence for the presence of implicit false belief throughout the lifespan (i.e., System 1). In their study based on the anticipatory looking paradigm, 3-year-olds, 4-years-olds, and adults were administered an identity false belief task and a location false belief task. In this study, the authors demonstrated that children and adults fail to correctly anticipate an actor's false belief about an object's identity based on visual perspective. Importantly, the 3-year-olds, 4-years-olds, and the adults correctly anticipated (i.e., when measuring first look) the actor's actions in the location false belief task. Thus, there is evidence for one of the signatures of System 1: efficient false belief tracking from childhood to adulthood, as measured with anticipatory looking.

To conclude, this study sheds light on the limitations of implicit false belief tasks based on anticipatory looking when used with populations across the lifespan. The current results illustrate the importance of using sound methodological practices (e.g. large samples) as well as a task that reliably measures implicit theory of mind through anticipatory looking. In other words, researchers should select a task that reliably tests implicit false belief understanding when using an anticipatory looking paradigm. Although the first look is more spontaneous and might better reflect automatic processing, our results seem to support DLS as a more reliable variable, at least when using the present instantiation of the anticipatory looking false belief task. For the adults, the DLS seems robust when higher task demands are involved, or when the task biases the participants towards one side of the screen (i.e., when the agent/car leaves from one side of the screen rather than from the center). The present results allow for a better understanding of the extent of participants' implicit false belief understanding. Future research will be required to fully examine the participants' performance on simple implicit false belief tasks by using several measures (i.e., first look and DLS) and by examining where the participants who fail the task are looking. It appears that a wide range of factors are responsible for the reported variability in performance across studies. We therefore encourage further work to examine children's and adults' performance on other anticipatory looking tasks measuring false belief, which may help shed light on the nature of implicit false belief.

Acknowledgements

This research was supported by research grants from the Social Sciences and Humanities Research Council of Canada (SSHRC; #435-2012-1403) and the National Institute of Child Health and Human Development (NICHD; #R01HD068458) to Diane Poulin-Dubois. The authors would like to thank Bethsheba Ananng for her assistance in testing.

References

Apperly, I. A., & Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review, 116*(4), 953–970. http://dx.doi.org/10.1037/a0016923.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a 'theory of mind'? *Cognition, 21*(1), 37–46. http://dx.doi.org/10.1016/0010-0277(85)90022-8.

Carruthers, P. (2016). Two systems for mindreading? *Review of Philosophy and Psychology, 7*(1), 141–162. http://dx.doi.org/10.1007/s13164-015-0259-y.

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

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G*power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*(4), 1149–1160. http://dx.doi.org/10.3758/BRM.41.4.1149.

Grosse Wiesmann, C., Friederici, A. D., Singer, T., & Steinbeis, N. (2016). Implicit and explicit false belief development in preschool children. *Developmental Science*. http://dx.doi.org/10.1111/desc.12445.

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*(3), 305–311. http://dx.doi.org/10.1177/0956797612451469.

Low, J., Apperly, I. A., Butterfill, S. A., & Rakoczy, H. (2016). Cognitive architecture of belief reasoning in children and adults: A primer on the two-systems account. *Child Development Perspectives, 10*(3), 184–189. http://dx.doi.org/10.1111/cdep.12183.

Low, J. (2010). Preschoolers' implicit and explicit false-belief understanding: Relations with complex syntactical mastery. *Child Development, 81*(2), 597–615. http://dx.doi.org/10.1111/j.1467-8624.2009.01418.x.

Perner, J., Leekam, S. R., & Wimmer, H. (1987). Three-year-olds' difficulty with false belief: The case for a conceptual deficit. *British Journal of Developmental Psychology, 5*(2), 125–137. http://dx.doi.org/10.1111/j.2044-835X.1987.tb01048.x.

Schuwerk, T., Vuori, M., & Sodian, B. (2015). Implicit and explicit theory of mind reasoning in autism spectrum disorders: The impact of experience. *Autism, 19*(4), 459–468. http://dx.doi.org/10.1177/1362361314526004.

Schuwerk, T., Jarvers, I., Vuori, M., & Sodian, B. (2016). Implicit mentalizing persists beyond early childhood and is profoundly impaired in children with autism spectrum condition. *Frontiers in Psychology, 7*, 1696. http://dx.doi.org/10.3389/fpsyg.2016.01696.

Senju, A., Southgate, V., White, S., & Frith, U. (2009). Mindblind eyes: An absence of spontaneous theory of mind in Asperger syndrome. *Science, 325*(5942), 883–885. http://dx.doi.org/10.1126/science.1176170.

Senju, A. (2012). Spontaneous theory of mind and its absence in autism spectrum disorders. *The Neuroscientist, 18*(2), 108–113. http://dx.doi.org/10.1177/1073858410397208.

Sodian, B., Licata, M., Kristen-Antonow, S., Paulus, M., Killen, M., & Woodward, A. (2016). Understanding of goals, beliefs, and desires predicts morally relevant theory of mind: A longitudinal investigation. *Child Development, 87*(4), 1221–1232. http://dx.doi.org/10.1111/cdev.12533.

Southgate, V., Senju, A., & Csibra, G. (2007). Action anticipation through attribution of false belief by 2-year-olds. *Psychological Science, 18*(7), 587–592. http://dx.doi.org/10.1111/j.1467-9280.2007.01944.x.

Thoermer, C., Sodian, B., Vuori, M., Perst, H., & Kristen, S. (2012). Continuity from an implicit to an explicit understanding of false belief from infancy to preschool age. *British Journal of Developmental Psychology, 30*(1), 172–187. http://dx.doi.org/10.1111/j.2044-835X.2011.02067.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*(2), 147–176. http://dx.doi.org/10.1111/mila.12099.

Wang, B., Hadi, N. S. A., & Low, J. (2015). Limits on efficient human mindreading: Convergence across Chinese adults and Semai children. *British Journal of Psychology, 106*(4), 724–740. http://dx.doi.org/10.1111/bjop.12121.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development, 72*(3), 655–684. http://dx.doi.org/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*(1), 103–128. http://dx.doi.org/10.1016/0010-0277(83)90004-5.