Is Implicit Theory of Mind the 'Real Deal'? The Own-Belief/True-Belief Default in Adults and Young Preschoolers

LU WANG AND ALAN M. LESLIE

Abstract: Recent studies reveal spontaneous implicit false-belief understanding in infancy. But is this early ability genuine theory-of-mind? Spontaneous tasks may allow early success by eliminating the selection-response bias thought to underlie later failure on standard (verbal) tasks. However, using anticipatory eye gaze, we find the same bias in non-verbal tasks in both preschoolers and adults. We argue that the bias arises from theory-of-mind competence itself and takes the form of a rational prior to attribute one's own belief to others. Our discussion then draws attention to a number of other inferential hallmarks of early belief-desire reasoning that together suggest it *is* the real deal.

1. Introduction

If we want to go to the museum in an unfamiliar town, it is natural to ask a passerby for directions. We assume, unless we have a specific reason to think otherwise, that the passerby's beliefs about the museum's location will be correct (and their expression sincere). If we did not assume this, then why ask in the first place? We make this assumption not only when we want information we currently lack, but also, we claim, whenever we already have the relevant information and are trying to infer the beliefs of another person. We will call this attribution, the true-belief default.

Dennett (1981) drew attention to a related aspect of the default when he pointed out that in order to do their job beliefs *ought* to be true. In the context of theory of mind, the true-belief default can simplify the computational task of calculating the beliefs of others, at least whenever we too have a belief about the same matter. It simplifies our computation by giving it a constrained starting point. Moreover, when it comes to commonplace, everyday matters, a person's beliefs usually *are* true, underwriting the general validity of taking our *own* belief as a best guess for another person's belief. In this guise, the true-belief default is a rational prior.

We thank Charles Randy Gallistel, Rochel Gelman, and Renee Baillargeon for their comments on earlier version of the manuscript. We also thank the members of Cognitive Development Lab at Rutgers University, and the schools, families, and undergraduate students at Rutgers who participated. Special thanks to Victoria Southgate for generously sharing her video used in the LD condition of experiment 2. This research was supported by National Science Foundation grant (BCS-0922184) to AML.

Address for correspondence: L. Wang, Department of Psychology and Center for Cognitive Science, Rutgers University, 152 Frelinghuysen Road, Piscataway, NJ 08854, USA.
Email: luwang@ruccs.rutgers.edu

Mind & Language, Vol. 31, No. 2 April 2016, pp. 147–176.

1.1 Calculating False Beliefs

Predicting the behavior of another person depends upon an ability to compute that person's perspective on the world. We do this by trying to determine a pair of contents, one for a 'belief', the other for a 'desire', that together allow the prediction of an action assuming an agent and an environment. The cognitive basis of this skill, or 'theory of mind', emerges in children early, before formal education begins. An important instrument in studying theory of mind in both typically developing children and its impairment in children with autistic spectrum disorders has been the 'Sally-Anne false belief' task (Baron-Cohen, Leslie and Frith, 1985). In the task scenario, one actor, Sally, hides a desired object in location A and then departs. A second actor, Anne, then moves this object hiding it again in location B. The child is asked to predict where Sally will search for the desired object upon her return. It is well-established that the majority of typically developing children aged four and a half years onwards pass this task by predicting that Sally will search location A where she still believes the object to be.

Based on such findings, most researchers reached a consensus that 'theory of mind' did not develop until four years or later when a conceptual deficit was overcome (Perner, 1995; Wellman, Cross and Watson, 2001). Some, however, argued that this consensus was premature (Bloom and German, 2000; Roth and Leslie, 1991; Scholl and Leslie, 2001). According to the latter view, the apparent failure of younger children was more likely due to the processing demands of standard false-belief tasks than to a conceptual deficit. For example, evidence was mounting that the standard false-belief task requires considerable executive resources. In the absence of contrary information, it is rational to attribute to others the belief one holds oneself because, as we remarked earlier, typically people's everyday beliefs are true (Leslie and Thaiss, 1992). However, in false belief tasks this default attribution is guaranteed to fail and, according to the Leslie and Polizzi (1998) model, must first be inhibited if the correct belief content is to be selected.

Although computational models of early belief-desire reasoning are rare, two algorithms have been outlined (Leslie and Polizzi, 1998) and investigated experimentally (Friedman and Leslie, 2004a, 2004b, 2005; Leslie, Friedman and German, 2004; Leslie, German and Polizzi, 2005; see also Baker, Saxe and Tenenbaum, 2009). According to these models, early belief-desire reasoning is a spontaneous process that infers first a content for an agent's belief b given the agent's environment, and then an action a given the agent's desire d and belief b in that environment. This process occurs in two steps, first a theory-of-mind mechanism (ToMM) provides plausible hypotheses for belief contents; typically among these hypotheses is a 'true-belief' (TB) that corresponds to the child's own belief about the scenario. Additionally, in a false-belief task, ToMM proposes a plausible content for the 'false-belief' (FB) typically based on the agent's visual (or other informational) access to the scenario. Second, a selection process (SP) reviews the hypotheses and, by default, selects the TB content (own belief) for attribution unless it has sufficient power to inhibit this default tendency, thus selecting the content based on history of exposure e, the FB content.

1.2 New Wave, Old Questions

The early competence theory has received support in recent years from a new wave of studies that have adapted standard false-belief tasks eliminating verbal content or reducing it to a minimum. These tasks also use indirect measures, such as violation-of-expectation looking time, anticipatory eye gaze, and spontaneous helping (e.g. Onishi and Baillargeon, 2005; Southgate, Senju and Csibra, 2007; Buttelmann, Carpenter and Tomasello, 2009; Kovács, Téglás and Endress, 2010; Scott, Baillargeon, Song and Leslie, 2010). Early success on false-belief tasks brings later failure into sharp relief. However, the older findings do not simply disappear with the new; rather, they highlight the need to understand the factors preventing successful performance at an explicit verbal level.

A plausible possibility is that early-success tasks do not make demands on immature executive processes thus revealing the early competence (cf. Apperly and Butterfill, 2009). One suggestion is that early-success tasks rely purely on spontaneous responding whereas later-failure tasks demand 'elicited' responses such as pointing or verbalizing (Baillargeon, Scott and He, 2010; Scott, Baillargeon, Song and Leslie, 2010). If spontaneous responding does not require executive processes, then we can explain both early success and later failure.

However, some violation-of-expectation looking time studies hint that executive demands may still be at work. For example, Scott *et al.* (2010, p. 383) found evidence that infants looked longer when viewing behavioral violations of false beliefs than when viewing behavioral violations of true beliefs, suggesting that false-belief processing is more challenging for infants, even in non-verbal spontaneous-response tasks. In another looking-time task, Yott and Poulin-Dubois (2012) found that executive function positively correlated with infants' looking at the event in which behavior violated the actor's false belief. The higher the infant scored in executive function tasks, the longer the infant looked at the unexpected outcome in the false-belief task, suggesting that the inhibitory demands of false belief may influence violation-of-expectation looking.

Looking-time methods are not the only source of evidence for precocious theory of mind. Anticipatory eye gaze, measured automatically by an eye tracker, shows that 24-month-olds can spontaneously predict the behavior of an actor with a false belief (Southgate *et al.*, 2007). Infants watched a videoed scenario modeled on the Sally-Anne task. An object is hidden in one container while an actor observes and then relocated, first to another container and then, while the actor's back is turned, removed entirely from the scene. Finally, the actor turns to face the scene once more and a chime signals that the actor is about to reach for the object. Eye-tracking measures reveal that infants' eye gaze spontaneously anticipates that the actor will search the container in which she last saw the object hidden (Southgate *et al.*, 2007).

Recent studies of adults using anticipatory eye gaze suggest that false-belief processing may be subject to executive demands (Schneider, Bayliss, Becker and Dux, 2012; Schneider, Lam, Bayliss and Dux, 2012; cf. Low and Watts, 2013). Schneider, Lam, Bayliss and Dux (2012) measured subjects' anticipatory looking in a false-belief task modeled on Sally-Anne in a dual task paradigm. Adults' spontaneous belief reasoning appeared to be abolished when a concurrent task was introduced, leading these authors to argue that spontaneous belief reasoning requires executive processes.

1.3 What Drives Performance Limitations?

In an effort to better understand the key differences in the processing demands of standard verbal false-belief tasks and early-success non-verbal tasks, we designed two experiments using anticipatory eye-gaze measures. Existing studies using this measure have tested either adults using only 'high-demand' tasks (the target object remains hidden on scene and subjects know its position), or preschoolers using only 'low-demand' tasks (subjects are uncertain about the target object's whereabouts after the object is entirely removed from the scene). Removing the target object off-scene to an unknown location should undermine the subject's certainty about the true belief, making it easier to attribute a false belief to the actor.

Several existing studies support the above idea. For example, Zaitchik (1991) uses an untrustworthy actor to inform the child about the object's location without the child being able to see for herself where it is. In Bartsch (1996), the actor announces where she thinks the object is but the child sees that both locations are in fact empty. In Carpenter, Call, and Tomasello (2002), an actor points at a box in which she previously placed a (novel) object—but which, unbeknownst to the actor, had subsequently been removed—and says, 'Let's get the dax!'. In Kikuno, Mitchell and Ziegler (2007), an actor has a Plasticine hat, and while he is away, the Plasticine is reshaped to an apple before being relocated to another location. Children are asked where the main character would look for his hat.

In all of these cases, the child herself either does not have a competing belief about the specific location of the target object (Bartsch 1996; Carpenter *et al.*, 2002; Kikuno *et al.*, 2007) or has less confidence in that belief, derived, as it is, from an untrustworthy informant (Zaitchik, 1991; also Roth and Leslie, 1991). Less competition from the subject's own belief will make it easier to attribute an alternative belief to the actor, creating a 'low-demand' false-belief task. In Southgate's anticipatory eye-gaze task, a puppet removes the target from the scene and disappears with it to an unknown location (thus 'low-demand'). In standard Sally-Anne tasks, by contrast, Anne moves the object from one hiding place to another in the same scene, so that the subject knows its current position (thus 'high-demand').

1.4 Goals of the Present Studies

Our first goal was to determine whether 'new wave' early-success tasks are free from executive demands or whether the same or similar demands arise in these tasks too. To date, no single study has contrasted a full complement of implicit tasks, namely, low-demand false belief, high-demand false belief, and true belief. Furthermore, the anticipatory eye gaze paradigm allows us to use the exact same tasks with both young preschoolers and adults; we can therefore isolate the effects of performance variables from lack of competence.

Our second goal was to address the nature of performance demands in the false-belief task. There are two main views in the literature about what drives performance in the false-belief task. One hypothesis is that subjects simply report or respond to reality (Kikuno *et al.*, 2007; Mitchell, 1994; Russell, Mauthner, Sharpe and Tidswell, 1991; Rubio-Fernández, 2013; Schneider, Lam, Bayliss and Dux, 2012; see also Carpenter, Call and Tomasello, 2002, for 'the pull of the real'; and Fodor, 1992, and Cassidy, 1998, for a heuristic failure to consider belief). Instead of calculating belief, subjects for a variety of reasons simply report reality. Failures in false-belief tasks, then, reflect a failure to *employ* an otherwise effective theory of mind.[1]

A contrasting approach explained false-belief failures as arising from the nature of theory of mind itself. Drawing on modularity theory, Leslie (1994) predicted that belief calculation would occur spontaneously and at the same time be subject to a 'true-belief bias', reflecting its status as a rational prior (Leslie, German and Polizzi, 2005; Leslie and Polizzi, 1998; Leslie and Thaiss, 1992).

Most of the time, the competing hypotheses of 'reality bias' and 'true-belief bias' predict the same response. If subjects either do not possess or do not use theory of mind, they must focus only on the state of the world and predict actors' behavior as a direct response to objects and their locations in the world. Specifically, in the standard Sally-Anne task a 'reality bias' predicts a focus on the current hiding location of the marble. At the same time, theory of mind with uninhibited 'true-belief' default, while positing very different computations and representations, namely, an attribution of belief to Sally, predicts the same answer in this task (and in many other experiments).

The difference between the reality and true-belief biases are subtle and there has been a paucity of data in the literature directly addressing the distinction. Indirect evidence comes from the standard task, which typically includes a 'memory' control question. Children are required to pass that question and most do, easily ignoring current reality. One previous study that did directly address this question used reaction time measures in conjunction with traditional questioning (Kikuno, Mitchell and Ziegler, 2007). Across three experiments, these researchers found clear evidence that false belief questions took preschoolers longer to answer than reality and memory questions. Furthermore, false belief questions took longer to answer than true belief questions—even when children got false belief wrong and gave the same answer to both questions. Crucially, Kikuno *et al.*'s (2007) experiment 3 also allowed comparison of true belief questions with reality questions. True belief questions took longer to answer than the corresponding reality questions, despite having the same answers; reasoning about beliefs, true or false, is different from answering questions about current and past realities.

[1] Conceptual deficit accounts of development (Gopnik and Wellman, 1995; Carlson, Moses and Breton, 2002; Perner, 1991) assume a lack of competence and thus entail that deficient subjects can report only 'reality'.

In two experiments, we explore, using implicit non-verbal scenarios and measures, (a) whether early success tasks evidence similar performance demands as traditional late success tasks, (b) whether such demands arise from reality or true-belief biases, and finally (c) whether there are differences between toddlers and adults on the same tasks.

2. Experiment 1

We adapted and extended the design of Southgate *et al.* (2007). In that study, a video shows an actor standing behind a screen with her head poking above it. There are two windows in the screen, one on either side of the actor. In front of and under each window, there is a container. A puppet places and hides a toy in one of the two containers. Subsequently, the two windows illuminate, accompanied by a chime; together, these signal that the actor is about to reach for the toy. Infants were familiarized with this video in two trials, shown on an eye tracker. To test spontaneous belief attribution, infants were then shown a video in which the puppet entirely removed the toy from the scene while the actor had her back turned. The actor then turned around to face the scene, and the anticipatory signal was given while infants' eye gaze was recorded. In our terms, this was a low-demand (LD) false belief condition because the target was removed from the scene.

Southgate generously shared her video with us, and we edited it to make videos for two further conditions, a true belief (TB) and a high-demand false belief (HD). In the TB video, as the puppet is relocating the toy into the other container, the actor turns back to face the scene, and sees the toy in its final position. In the HD video, the actor does not turn back until after the toy has been relocated and the puppet has left the scene empty handed. Therefore, in the HD condition, the actor should falsely believe that the toy is still in the original container but in fact, it is still on scene hidden in the other container.

The FB-LD condition replicates Southgate *et al.* (2007) exactly. Based on that study, we therefore predict longer looking to the window through which the actor with a false belief about the target would reach. (A null hypothesis conceptual-deficit theory predicts diffuse looking to either side.) In the TB condition, we predict the opposite pattern of looking to the side where the object actually is. The pattern of looking in FB-HD condition can inform us about underlying processes. A reality bias will pull attention to the same side as TB. If early success tasks are free of bias then this condition should pattern with FB-LD and attention should flow to the false belief side. An uninhibited true-belief bias would produce an intermediate pattern reflecting an unresolved selection process with both false and true belief hypotheses still in play.

Because we can define the windows and the containers as separate Areas of Interest (AOI), we can determine more precisely what the focus of subjects' attention is across scenarios. In both TB and FB-HD conditions, a reality bias predicts that the differential focus of looking should go to the box containing the hidden object. By contrast, theory of mind predicts differential attention to search (action intention) rather than to successful outcome (where the object will be found). In short, a dominating concern with reality should highlight boxes over windows.

2.1 Methods

Subjects were 87 (47 girls) two- to three-year-olds (mean age = 36.3 months, range = 24.2 − 47.9 months) and 172 (85 females) Rutgers University undergraduate students (mean age = 19.6 years, range = 17.0 − 28.1 years). An additional child failed to complete because of experimenter error. For all analyses of anticipatory gaze (with one exception, see footnote 2), out of the aforementioned totals, four children were excluded due to calibration error, two children and nine adults were excluded because the Tobii eye tracker recorded no looking behavior during the four seconds of anticipation.

Subjects watched three videos showing two main familiarization trials, and a single test trial varying by condition. The videos for the familiarization trials and the low-demand false belief trial were the same as used in Southgate *et al.* (2007). The start scenes of all the videos were the same: an actor wearing a visor is shown standing behind a screen with her head poking above it. In the screen are set two square-shaped windows, one on either side of the actor. In front of and under each window, there is a cube-shaped container; the lid of each container is below the bottom of the window on each side.

In the test trial of LD false-belief condition, the puppet first puts the ball in the left-side container and leaves. A phone starts to ring at this point, and the actor turns her head around as if she is distracted and attending to the ring. While the actor has her back turned, the puppet comes up again, opens the lid of the left-side container, takes the ball out, opens and places the ball into the right-side container. After he closes both lids, the puppet pauses for a second in the center, and then opens the lid of the right-side container again, takes out the ball, closes the lid, and finally takes the ball entirely off the scene. The phone ring stops immediately after the puppet disappears, and the actor turns back to face the scene. The windows light up and the chime sounds, indicating that the actor is about to search for the ball. The windows light for 2 seconds after which the image freezes for a further 1.9 seconds. Unlike Southgate *et al.* (2007), no reaching outcomes were ever shown to subjects.

The HD false-belief condition was obtained by editing the above video, so that the puppet leaves the scene empty-handed after he relocated the ball from the left-side container into the right-side one. The actor now turns back around as before and the rest of the video is the same. (Loss of a few frames in editing meant the HD video ended with a 1.85 s frozen image.) The video for the TB condition was also obtained by editing the LD video, so that the actor turns back to face the scene while the puppet was putting the ball into the right-side container, and therefore saw the ball's final position. Because of editing demands, in the TB video, the windows stay lit for 2.8 s and the image is then frozen for a further 1.2 s. The container in which the actor last saw the ball being placed was counterbalanced across subjects and conditions.

A Tobii T60 XL eye tracker automatically made a 60 Hz record of the latency of subjects' first saccade and the duration of their gaze towards each of five areas of interest: both windows, both containers, and the head of the actor. After the four-second anticipation recording period, adults were asked, 'Where will she search for the ball?' Their responses were recorded as Left-container or Right-container from their own perspective, and further coded as zero (inconsistent with the actor's belief) or one (consistent with the actor's belief).

2.2 Results

First, we wanted to know whether subjects looked mainly to windows or to containers. Southgate *et al.* (2007) presented data only for windows.

2.2.1 Preferential Looking at Windows versus Containers. In general, subjects were interested in the agent and the windows, and paid relatively little attention to the containers. Specifically, subjects looked significantly longer at the windows than at the containers: children: $t\ (86) = 11.82$, $p < .001$, two-tailed; and adults: $t\ (171) = 12.01$, $p < .001$ two-tailed; see Figure 1. The preference for windows held no matter whether the target object was hidden on the scene as in TB or HD conditions: children, $t\ (53) = 9.29$, $p < .001$, two-tailed; adults, $t\ (105) = 9.62$, $p < .001$, two-tailed; or completely removed from the scene as in LD: children, $t\ (32) = 7.20$, $p < .001$, two-tailed; adults, $t\ (65) = 7.41$, $p < .001$, two-tailed.

To examine whether looking at containers was biased by the hidden object ('reality bias'), we combined subjects' looking times to the containers in the HD and TB conditions (target object hidden on scene), and compared that to subjects' looking times to the containers in the LD condition (both containers empty). Subjects did not look longer to the containers when the target was hidden on scene: children, $t\ (85) = 0.70$, $p = .49$, two-tailed, Bayes Factor = 4.7:1, odds in favor of the null; adults, $t\ (170) = 1.09$, $p = .278$, two-tailed, Bayes Factor = 4.6:1, odds in favor of the null. Next, we combined subjects' looking times to the 'reality' container in the HD and TB conditions, and compared that to the combined looking times to the empty container in these conditions. Subjects looked at each equally: children, $t\ (53) = 0.97$, $p = .34$, two-tailed, Bayes Factor = 5.9:1, odds in favor of the null; adults, $t\ (105) = 0.96$, $p = .34$, two-tailed, Bayes Factor = 8.3:1, odds in favor of the null. The Bayes Factor analyses have the advantage over traditional significance testing of being able to evaluate the extent to which the data support the null hypothesis and not just the experimental hypothesis.[2] These data provide substantial evidence that there was no tendency to look at the location of the hidden object, contradicting the 'reality bias' hypothesis.

[2] The resulting Bayes Factor is expressed as an odd ratio; although the odds are 'what they are' (Gallistel, 2009) an accepted rule of thumb for guidance is that odds equal to or greater than 3:1 provide 'substantial evidence' for a given hypothesis.

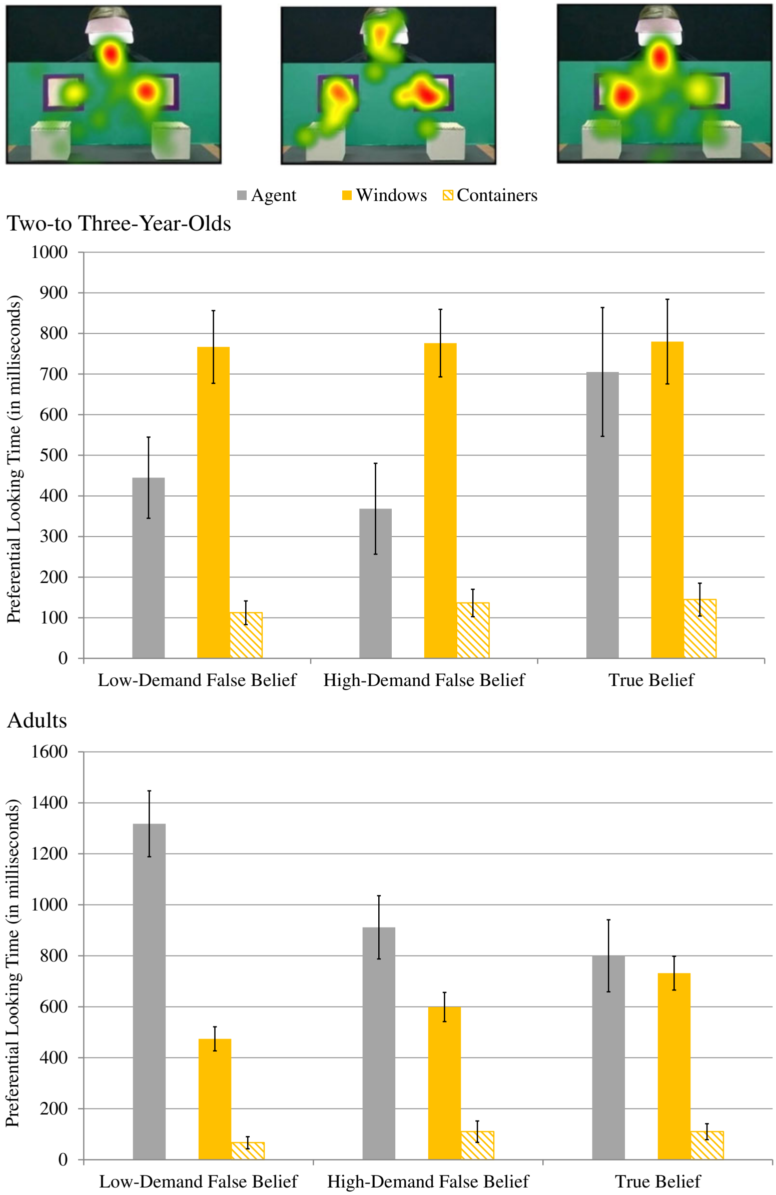

Figure 1 *Experiment 1, Looking to agent, windows and containers during 4 s anticipation test trial. Balance of looking was outside AOI or not recorded.*

2.2.2 Anticipatory Looking to Windows by Condition. Looking at the actor's visored face cannot reveal which of her behaviors subjects anticipate. Critically, the data presented above shows that looking at the containers is also uninformative. Therefore, we analyzed subjects' looking times to the windows as a measure of subjects' anticipation of the actor's behavior (Southgate *et al.*, 2007). Data from subjects who did not pay sufficient attention to either window were excluded from further analysis. Specifically, we excluded, for children, data from the bottom 5th percentile of total looking time to the windows in each condition (cut-off looking time, 233 ms in TB, 1 child; 33 ms in LD, 3 children; and 78 ms in HD, 1 child); and, for adults, the bottom 25th percentile (cut-off looking time, 479 ms in LD, 17 adults; 654 ms in HD, 16 adults; and 783 ms in TB, 13 adults).[3]

Subjects may shift first look following the chime to the window through which the actor should reach given her false belief ('FB window') or to the window through which the actor should reach given a true belief ('TB window'). Both children's and adults' anticipatory first looks were significantly different across conditions (Conditions (3) × First looks (2) two-tailed chi-square analysis, $\chi^2 = 9.61$, $p < .01$ for children, and $\chi^2 = 7.26$, $p < .05$ for adults). Given previous findings that children and adults made belief-consistent anticipations as measured by eye gaze in true-belief and low-demand false belief conditions (e.g. Schneider, Bayliss, Becker and Dux, 2012; Schneider, Lam, Bayliss and Dux, 2012; Senju, Southgate, White and Frith, 2009; Southgate, Senju and Csibra, 2007), we conducted a set of planned comparisons on anticipatory first looks in each condition. In the LD condition, both two- to three-year-olds and adults were more likely to look first to the FB window (19/27 children and 28/44 adults; one-tailed binomial tests, $p = .026$, and $p = .048$, respectively). The opposite looking pattern was observed in the TB condition, and subjects preferred the TB window first (15/20 children and 21/31 adults; one-tailed binomial tests, $p = .021$, and $p = .036$, respectively). In the HD condition, however, subjects showed no first look preference: 16/29 children and 20/42 adults looked at the FB window first (one-tailed binomial tests, $p = .36$, and $p = .44$, respectively; for children, Bayes Factor = 3.8:1 in favor of the null and for adults, Bayes Factor = 5.0:1 in favor of the null). Combining first looks across ages, the LD and HD conditions were significantly different in the expected direction (Upton's $\chi^2 = 3.48$, $p = .031$, one-tailed) as were the HD and TB conditions (Upton's $\chi^2 = 5.49$, $p = .01$, one-tailed).

Because no reaching outcomes were ever shown in test trials, preferential looking time over the four seconds was a further indication of spontaneous anticipation. We calculated a differential looking score (DLS) by subtracting subjects' looking time towards the TB window from that to the FB window and dividing by total looking

[3] These criteria excluded 21% of subjects in total. Such a rejection rate is common in the literature using an automatic eye tracker. For example, Southgate *et al.* (2007) excluded 16 out of 36 (44%) two-year-olds using a different criterion based on first looks.

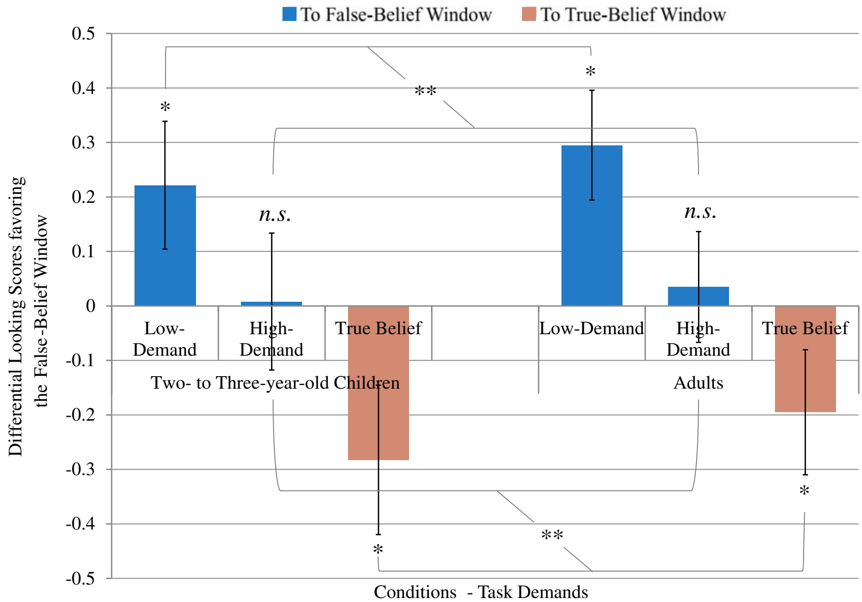

Figure 2 *Experiment 1, Differential Looking Scores (DLS) by Condition and Age. A positive DLS represents a preferential looking at the False-Belief Window calculated over the four seconds of anticipation. A negative DLS represents a preference for the True-Belief Window over the same period. ** = t-test, p < .05, two-tailed; * = one-sample t-test, p < .05, one-tailed.*

time to both windows. Therefore, DLS varied from −1 to +1, a positive number indicating a preference for the FB window (see Figure 2). Preliminary analyses showed that there were no main effects of children's age (two versus three years), gender, or initial location of the toy on DLS and no interactions with conditions, and likewise no main or interaction effects for adults. Therefore, we dropped these factors from further analysis.

ANOVA with Age (2) (preschoolers, adults) × Conditions (3) showed only a main effect of condition (F (2,187) = 8.29, $p < .001$, $\eta^2 = .081$). No other effects or interactions reached significance. Planned t-tests showed that mean DLS were significantly different between LD and HD (t (140) = 2.22, $p = .014$, one-tailed) and between HD and TB conditions (t (120) = 2.14, $p = .017$, one-tailed). In the LD condition, subjects showed a preference for the FB window with DLS = 0.22 for children (t (26) = 1.89, $p = .035$, one-tailed), and DLS = 0.30 for adults (t (43) = 2.93, $p = .003$, one-tailed)). In the TB condition, subjects showed a preference for the TB window with DLS = −0.28 for children (t (19) = 2.06, $p = .027$, one-tailed) and DLS = −0.20 for adults (t (30) = 1.70, $p < .05$, one-tailed). In the HD condition, subjects spent equal amounts of time looking at the TB and FB windows (mean DLS = 0.01 for children (t (28) = 0.06, $p = .475$, one-tailed, Bayes Factor = 6.96:1, odds favoring the null), mean DLS = 0.03 for adults

(t (41) = 0.34, $p = .367$, one-tailed, Bayes Factor = 7.85:1, odds favoring the null)).

2.2.3 Verbal Responses (Adults). Adult subjects were asked, 'Where will she look for the ball?' immediately after the videos. Results are reported for all adult subjects (172), and for those who were included for the anticipatory looking analysis (117), separately. For all adult subjects, 60/66 in LD, 57/62 in HD, and 38/44 in TB correctly predicted the actor's behavior given her belief (two-tailed binomial tests, all *P's* < .001). Their verbal responses were more accurate than their spontaneous anticipation by first look in all conditions (McNemar Binomial tests, two-tailed: in LD, $p < .001$; in HD, $p < .001$; in TB, $p = .013$). For those who were included in the analysis of looking at the windows, 42/44 in LD, 38/42 in HD, and 26/31 in TB correctly predicted the actor's behavior based on her belief (two-tailed binomial tests, all *P's* < .001). Their verbal responses were more accurate than their spontaneous anticipation by first look in both LD and HD conditions (McNemar Binomial tests two-tailed: in LD, $p = .001$; in HD, $p < .001$; in TB, $p = .227$).

2.3 Discussion

Two- to three-year-olds and adults correctly attributed a belief to an actor in our true-belief scenarios, but also in our false-belief scenarios when processing demands were low. However, knowing the hidden target's specific whereabouts reduced to 'chance levels' subjects' ability to anticipate the actor's behavior in high-demand false-belief scenarios. Strikingly, these demands similarly reduced the ability of our adult subjects. Together, these findings suggest that spontaneous non-verbal anticipation—both first look and preferential anticipation—does not escape the executive demands of non-verbal false-belief tasks when measured by anticipatory eye gaze.

2.3.1 Triple Pattern. False-belief tasks, regardless of the number of conditions, almost always entail one of only two response patterns, 'pass' or 'fail'. The present study, however, produced three distinct patterns. In the TB condition, eye-gaze shifted rapidly to the window through which the actor would reach towards the container that she and the subject knew contained the desired object. In the FB (LD) condition, eye-gaze shifted rapidly to the opposite window through which the actor would reach the container that she believed contained the desired object but which the subject knew was empty. The FB (HD) condition was statistically distinct from both of these. In this case, eye-gaze was equally likely to shift to either window, as if the true/own belief and the false belief drew subjects' attention equally.

A similar result is discernible from Schneider, Bayliss *et al.* (2012)'s dual-task eye tracking study of spontaneous attention in adults. Subjects, watching nonverbal videos modeled on the Sally-and-Anne story, looked more at the empty location in the false-belief condition than in the true-belief condition (when there was no extraneous task). However, within the false-belief condition, subjects looked equally long at both the empty and full locations, echoing our finding in experiment 1. Unfortunately, this study examined only what we are calling a high demand false-belief task and did not include a low-demand condition, so we are unable to say whether their adults would have shown a triple pattern.

Heyes (2014a, 2014b) explained toddlers' success in the study by Southgate *et al.* (2007) as driven by low-level perceptual features of the stimuli. In particular, Heyes (2014a, p. 654) suggested that, 'the bell ringing and head turning that was supposed to signal to the infants that the agent could not see movements of the ball, may instead have distracted the infants so that they didn't see, or didn't remember, those movements'. The 'distraction' of the agent turning away seems an unlikely explanation for our pattern of results. Our adults showed the same pattern of differential looking across conditions as our toddlers and yet later demonstrated an unimpaired memory for the basic movements of the scenarios. In any case, subjects in both age groups, prior to test, show very similar patterns of gaze to the puppet, object, and containers, and indeed to the agent herself, whether she has turned away or not.

Finally, we can rule out immaturity as a cause of the triple pattern in experiment 1, given that our adult subjects behaved the same way as our toddlers. Because of the novelty of the triple pattern finding, it is important that we test for this again with different stimuli. This is a major goal for experiment 2.

2.3.2 Reality Bias or Own-Belief/True-Belief Bias?. In many situations, the competing hypotheses of 'reality bias' and 'true-belief bias' will predict the same response. However, in another sense, the two possibilities could hardly be more different. In the latter case, the subject is employing theory of mind; in the former, she is not. The rational for the two reality-bias hypotheses we have considered is that either the 'pull of the real' temporarily disables theory-of-mind processing or a conceptual deficit means the subject has little option but to focus on world-behavior interaction. Either way, the pull of the real in both true-belief and high-demand false-belief conditions predicts that the actual location of the hidden object will draw eye gaze to the boxes. At the least, a 'pull of the real' should predict greater looking to the boxes when the object is hidden in one of them compared to when it is not, as in low-demand false belief. Neither of these predictions are supported by our findings.

However, though our findings contradict a reality bias, they do not yet support the true-belief bias. Subjects' attention may be drawn away from the containers simply because the windows illuminate or because they learned from familiarization trials that 'chime, now something happens around the windows.' Though low-level explanation does not account for the differential attention to the two windows across conditions, it may explain why subjects do not look at boxes at that juncture. The next experiment rules out such possibilities while testing the same bias predictions.

3. Experiment 2

In experiment 2, we introduced a novel split-screen design to measure anticipatory eye gaze. Subjects watched familiarization videos featuring a toy bunny and carrot for two trials. An actor puts the bunny in one of two boxes then says, 'Let me get its carrot'. She turns around to fetch the carrot, turns back, and places the carrot on the table between the boxes. At this point, the video splits into two smaller pictures that replace the original, one on the left and one on the right. These two pictures or 'screens', as we shall refer to them, are identical, except that, in the left screen, the actor is reaching to the left box, and, in the right screen, the actor is reaching to the right box (see Figure 3). The second familiarization video also has this structure but is counterbalanced, for side of initial placement of the bunny.

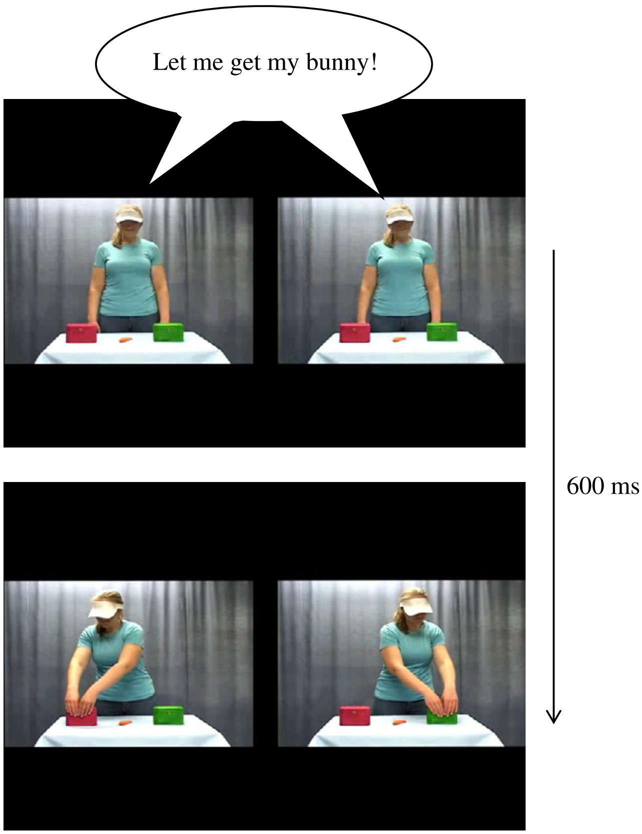

Figure 3 *Scenes following the splitting of the familiarization videos. Immediately following the split a central soundtrack carries the voice of the actor lip-synced with both screens; the actor then reaches to the red box in the right screen while in the left screen she reaches to the green box. The entire portion runs for 600 ms.*

The splitting of the video into two simultaneous versions with different endings serves two main purposes. First, the two screens allow measurement of differential action anticipation—the left screen always depicts the actor reaching to the box on the left, and the right screen always depicts the actor reaching to the box on the right. Second, the splitting takes the place of the chime-plus-window-illumination as the signal that the actor is about to perform an action. We reasoned that infants would learn in familiarization trials that the video would split and that the left screen would show the actor reaching to the left box while the right screen would show the actor reaching to the right box. In this way, infants in the test trials would be able to demonstrate action anticipation by shifting their gaze to whichever screen depicted the action they anticipated. In no case did we show the actor actually opening a box.

Following familiarization, subjects watched test videos featuring the same actor but this time the actor had a toy dog and a bone. The videos began the same way as previously but then, with the actor's back turned, differed according to experimental condition:

> *False-belief (FB) High-demand (HD) condition*, the dog crawled out of the original box and goes into the other box.
> *FB Low-demand (LD) condition,* the dog comes out of the first box and goes into the other box but then comes out of the second box, moves to the center of the stage, and leaves the scene entirely by jumping off the stage center front.
> *True-belief (TB) condition,* the dog moves from the original box to the other box, but the actor turns back to face the scene while the dog is climbing into the new box, and therefore, sees the dog's final position.

At these junctures, the video picture split into two identical screens, right and left, the actor says, 'Let me get my dog', and reaches right in the right screen and left in the left screen. An automatic eye-tracker recorded subjects' eye-gaze fixations for a period of 600 ms following the split.

The aims of experiment 2 are twofold. First, we want to see if subjects' spontaneous gaze anticipates that the actor's search will be in the appropriate opposite locations in the LD and TB conditions while in the HD condition anticipatory eye-gaze is again evenly balanced between the two. This last prediction requires testing the truth of a null hypothesis, for which we will use Bayesian Factor analysis (Gallistel, 2009).

Second, following the split, the real location of the object is the same in both screens. All varieties of a reality bias therefore predict the same kind of looking in both FB-HD and TB conditions. If the real directly 'pulls' the subjects' attention, looking will be split evenly between both screens (and, on a conceptual deficit version, so too in FB-LD); if instead the real 'pulls' the actor's reach, then subjects should look to the same screen in both FB-HD and TB conditions with an even split between both screens in FB-LD. By contrast, a true-belief bias predicts anticipation of the actor's reach to the belief-appropriate screen in TB, of actor's reach to the belief-appropriate screen in FB-LD, and evenly balanced looking between screens in FB-HD.

3.1 Methods

Subjects were 87 (45 girls) two- to three-year-olds (mean age = 37.0 months, range = 24.0 − 47.8 months, SD = 7.3 months) and 113 (54 females) undergraduate students from Rutgers University (mean age = 19.6 years, range = 18.0 − 28.3 years, SD = 1.5 years). About the same number of subjects in each age group were randomly assigned to Conditions: FB HD: 27 preschoolers (mean age = 37 m) and 35 adults (mean age = 19.2 yrs), FB LD: 30 preschoolers (mean age = 36 m) and 36 adults (mean age = 19.6 yrs), and TB: (30 preschoolers (mean age = 38 m) and 42 adults (mean age = 19.9 yrs). An additional 19 children and 10 adults were excluded because no looking behavior during the anticipatory phase was recordable (6 children, 10 adults) or the quality of looking was poor (overall looking time towards both screens < 200 milliseconds (13 children)).

Subjects watched two familiarization videos and a test video. Each video showed an actor in blue shirt behind a rectangular table, wearing a beige visor that covered her eyes. There were two wooden boxes, red and green, on opposite sides of the table. At the beginning of the first familiarization trial, the actor puts a toy bunny into the red box, closes the lids of both boxes, and says, 'Let me get its carrot', and turns away. She turns back with a toy carrot, says, 'Here it is', and puts the carrot down in the center of the table. The main picture then splits into two identical screens, each a copy of the main picture that they replace. Each screen was half the main picture's height and width presented against a black background. At 80 cm, the visual angle of the main picture before splitting was 30.2°; the visual angle of each split-screen was 15.5°, and the visual angle for the gap between the two split screens was 1.5°. The actor stands in the center of each screen, and says, 'Let me get the bunny.' After 600 milliseconds, in the screen on the left, she reaches to the left and grasps the lid of the red box, and in the screen on the right, she reaches to the right and grasps the lid of the green box. Her actions in both screens mirror each other and happen simultaneously. As soon as the actor touches the lid of a box, the action freezes until the end of the trial, seven seconds after the initial split. The second familiarization trial was identical to the first, except that the actor placed the bunny in the green box. The familiarization trials allowed subjects to discover that the main picture splits into two screens before the actor reached, and that the two screens would show two different outcomes simultaneously, with the left-hand reach occurring in the left screen and the right-hand reach occurring in the right screen.

In test trials, the actor puts a dog puppet into the red box on the left, and says, 'Let me get its bone', and turns away. The video then varies depending on the test conditions (see above). In false belief conditions, the location of the dog changes while the actor's back it turned. In FB-HD, the dog crawls out of one box and moves into the other, then remains hidden on scene. In FB-LD, the dog again crawls out of one box and moves into the other, but then comes out again, moves to the center of the table and jumps off exiting the scene entirely. In the TB condition, the actor turns back while the dog is climbing into the other box so that she sees where it ends up. In all conditions, the actor holding a blue bone in her hand, says, 'Here it is', and puts the bone down on the center of the table. After the screen splits, the actor says, "Let me get my dog," and reaches to a box, rightward in the right screen and leftward in the left screen.

Stimuli were presented on a Tobii (Stockholm, Sweden) T60 XL Eye Tracker, which also recorded subjects' eye gaze. A computer with Tobii studio presentation software controlled the study and recorded gaze. Adult subjects were tested individually in the lab. Children either came with their parents to the lab or participated at their preschool. When tested at preschool, a portable T60 was used to test children in a separate quiet room. Children sat 60 to 80 cm from the eye tracker (individually calibrated to maximize tracking quality). When tested in the lab, children sat in a testing booth on their parent's lap, and were also individually calibrated at 60 to 80 cm from the screen. Adult subjects sat on a chair in the same environment. Children's and adults' eye tracking was calibrated with age-appropriate 5-point Tobii calibration videos at the start of the experiment.

Each subject watched three videos of two familiarization trials followed by a single test trial. On test trials, we measured subjects' anticipatory first look (gaze shift) to one or other screen following the first video frame that showed the split-screens. Only gaze within the first 600 milliseconds following the split (while the two screens were still identical, before reaching outcomes began) was considered anticipatory. Immediately after the test trial, adults were asked, 'Where will she look for her dog?' Their responses were recorded as Red box or Green box and further coded as 0 (inconsistent with the actor's belief) or 1 (consistent with the actor's belief).

3.2 Results

The split-screen method shows both possible reaching outcomes in all conditions. For example, in the HD condition, one split-screen will show the actor reaching to the box consistent with her outdated belief (the false-belief (FB) screen). The other split-screen will show an outcome consistent with the subjects' own belief (the true-belief (TB) screen). Given that the two screens were identical during the first 600 milliseconds, and that the actor always reached to the left in the left-side screen and to the right in the right-side screen, any systematic preference for shifting first look to one side could only be explained by a preference for the outcome that would be shown in that screen. Therefore, preference for a particular outcome reflects subjects' anticipation of the actor's behavior prior to the splitting. The question of interest was whether the processing demands of the test trial would influence subjects' first-look anticipation during the 600 milliseconds.

Figure 4 shows the percentage of first looks to each screen by condition and age group. Both children's and adults' anticipatory first looks were significantly different across conditions (Conditions (3) × First looks (2) two-tailed Chi-square analysis, Upton's $\chi^2 = 9.84$, $p < .01$ for children, and Upton's $\chi^2 = 13.50$, $p < .01$ for adults). Following the results of experiment 1, we conducted a set of planned comparisons on anticipatory first looks. In the LD condition, both two- to three-year-old children and adults preferred to look first to the FB screen (21 out of 30 two- to three-year-olds, and 27 out of 36 adult subjects; one-tailed binomial tests, $p = .022$,

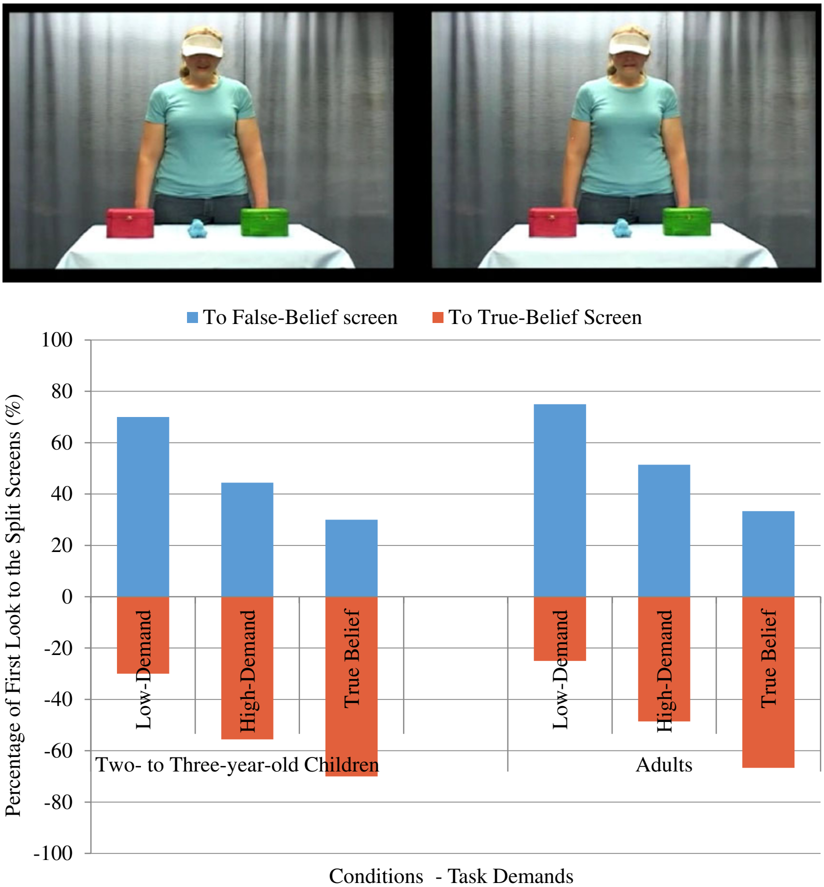

Figure 4 *Experiment 2, Anticipatory First Look. Percentage of two-to three-year-old children (left cluster) and adults (right cluster) whose first look was toward the False-belief screen (blue bar) and True-belief screen (red) in each condition.* * *= Binomial test,* $p < .05$, *one-tailed.*

and $p = .002$, respectively; see Figure 3). The opposite preference was observed in the TB condition in both children and adults (21 of 30 two- to three-year-olds, and 28 of 42 adult subjects; one-tailed binomial tests, $p = .022$, and $p = .022$, respectively). However, subjects showed no preference for first look to the FB screen in the HD condition (12 out of 27 two- to three-year-olds, and 18 out of 35 adult subjects; one-tailed binomial tests, $p = .35$, and $p = .50$, respectively; Bayes Factor analysis supports the null hypothesis of no difference with odds of 3.6:1 for children and 4.8:1 for adults). Given that children and adults showed the same pattern of anticipation, we combined the results across age groups. Analysis of combined anticipations show that first looks in LD and HD conditions were significantly different in the expected direction (Upton's $\chi^2 = 7.89$, $p = .002$, one-tailed); likewise the HD and TB conditions (Upton's $\chi^2 = 3.74$, $p = .027$, one-tailed). Critically, the evidence is 'strong' (odds > 10:1) that subjects' anticipatory first looks favored the TB screen in the TB condition (combined across age, 49 out of 72; BF = 16.6:1 odds favoring H_1). There was also 'substantial' evidence (odds > 3:1) that the first looks were balanced evenly between screens in the HD condition (30 out of 62; BF = 6.2:1 odds favoring H_0).

Immediately after the videos, adult subjects were asked, 'Where will she look for the dog?' They correctly predicted the actor's behavior given her belief in all three conditions (30 of 36 in LD, 31 of 35 in HD, and 37 of 42 in TB; all two-tailed binomial tests, $ps < .001$). Verbal responses were more accurate than spontaneous anticipatory gaze in HD and TB conditions but not in LD (McNemar Binomial tests: in HD, $p = .001$; in TB, $p = .022$; in LD, $p > .5$).

3.3 Discussion

We replicated the findings of experiment 1. Spontaneous, non-verbal anticipation measured by anticipatory eye gaze is subject to the varying executive demands of false belief tasks. We found again three distinct patterns of response, one for each of the conditions. The response pattern in the high-demand false-belief condition was again balanced equally between the tendency to attribute a false belief and the tendency to attribute a true belief. Low-level explanations of this triple pattern seem unlikely, for example, a first-look preference for container color or for meaningless motions would not produce these systematic preferences.

3.3.1 Anticipation of Action and Belief Attribution. Given that anticipatory saccades often take 200 ms to plan and launch (e.g. Altmann, 2011), can we really argue that (all of) our 600 ms window is anticipatory? We assume, in accord with the ToMM theory, that subjects track the agent's belief spontaneously. Support for this assumption comes from Kovács, Teglas and Endress (2010) who showed that both infants and adults spontaneously track an agent's belief, without external prompts. Thus, subjects may already have shifted attention covertly and begun saccade planning to one side or another even before the screen split. In this case, the split functions not so much to trigger belief attribution, than to give us a standard point from which to measure a response that has already taken place, albeit covertly.

A different question asks, is a 600 ms window long enough to reflect belief attribution? We measured fast, immediate, anticipatory first look. However, especially in the case of high demand false belief, perhaps subjects require a longer period to select the false belief as they struggle to inhibit their own (true) belief. A recent study by Scott, He, Baillargeon, and Cummins (2012), using a double picture array accompanied by a verbal narrative, measured preferential looking over a six-second period that followed a prompt plus a two-second 'inspection' period. They found that 2.5-year-olds preferred to look at the outcome matching the agent's false belief (which we would class as 'high-demand'). Note that Scott *et al.*'s study did not examine the effect of demand and measured a slower, non-anticipatory preference. Nevertheless, our subjects might have been able to anticipate correctly the actor's behavior in the high-demand false-belief condition, if only they had enough time. (Note that although we had four seconds for anticipation in experiment 1 without establishing a preference, a yet longer period might allow a preference to emerge.)

3.3.2 Reality Bias or Own-Belief/True-Belief Bias?. In both FB HD and TB conditions, the true location of the target object is in the same place on both screens after the split. As noted earlier, a reality bias predicts the same pattern of attraction across these conditions. This was not what we found. Instead, we observed again the same triple pattern found in experiment 1, this time without the effect of chimes and windows. We interpret these results as support for spontaneous belief attribution biased by a true-belief default.

4. General Discussion

It used to be widely assumed that false-belief reasoning first developed only some time after the fourth birthday (e.g. Perner, 1991; Wellman, Cross, Watson, 2001). More recent findings have revealed spontaneous theory of mind in the second year of life (or earlier) (e.g. Buttelmann *et al.*, 2009; Kovács, Téglás and Endress, 2010; Luo, 2011; Onishi and Baillargeon, 2005; Scott *et al.*, 2010; Southgate, *et al.* 2007; Southgate, Chevallier and Csibra, 2010; Surian, Caldi and Sperber, 2007; Träuble, Marinovic and Plauen, 2010). Successful spontaneous belief-desire reasoning has also been reported in toddlers and young preschoolers (Clements and Perner, 1994; Garnham and Ruffman, 2001; He, Bolz and Baillargeon, 2011; Scott *et al.*, 2012). The present findings add further support for early belief-desire reasoning.

Whereas the above studies and others support the early competence view of theory of mind (Carruthers, 2013; Scholl and Leslie, 2001; Roth and Leslie, 1991), there remains a wealth of findings from tasks showing later failure. Early competence theories explain later failure in terms of processing (performance) demands that the child is not yet able to meet (Bloom and German, 2000; Leslie, 1994; Leslie and Thaiss, 1992). Given this explanation, it is natural to assume that the early success tasks do not make these demands (Baillargeon, Scott and He, 2010). The present studies addressed this hypothesis directly and examined the long-term continuity of spontaneous early success measures by using the exact same tasks with preschoolers and adults.

We varied the processing demands of belief reasoning tasks in experiments 1 and 2 and measured two-to three-year-old children's spontaneous gaze anticipations. As predicted by early competence theories, the appropriate and opposite gaze patterns were observed in the low-demand false-belief and true-belief scenarios. Two- to three-year-old children expected an actor to behave in accord with her false belief when the actor believed that a now-absent target object was still hidden in one of the containers. When the actor saw the target object going into its last position, however, children anticipated the actor would behave consistently with a true belief. In the high-demand false-belief scenario in which, unbeknownst to the actor, the target object is relocated from one container to the other, children showed a distinct *third* pattern of eye gaze by anticipating that the actor would search equally in each container. This pattern of anticipation did not reflect a preschooler's lack of competence. Adults showed the same patterns, and did so even though they could correctly explain the actor's behavior afterward. Adults' verbal performance was less than perfect, with 9%–16% of subjects answering prediction incorrectly. We did not warn or instruct our adult subjects beforehand about the later questioning. This error rate then may reflect the difficulty of verbally encoding from memory. Indeed, O'Grady, Kliesch, Smith and Scott-Phillips (2015) suggest that explicit questions following belief-inducing stories presented implicitly yields a most challenging combination. On a modular view of the ToMM (Leslie, 1992), verbalized of theory of mind reasoning will occur external to, later than, and be more effortful than the spontaneous processing characteristic of the theory of mind module. Cohen and German (2009) show that spontaneously encoded belief information decays rapidly in the absence of a top-down effort.

4.1 Informational Access, True Belief and Indecision

As we outlined in the Introduction, there are theoretical reasons for supposing that *any* successful belief-desire reasoning system will have a 'true-belief' bias. Typically, most of our mundane beliefs about everyday matters are true; if they were typically false, we would quickly be in trouble. Consequently, if I have no specific information regarding the content of your belief about X, then my best guess about your belief (about X) will always be my own belief (about X). This means that, for any belief-desire reasoning system, attributing a 'true belief' *by default* is both computationally efficient and a rational prior. According to this approach, then, the 'true belief' bias together with the problem of how to qualify that assumption in special cases like false belief should be hallmark signature of theory of mind competence itself. (For related approaches in the neuroscience literature to the problem of the co-ordination of self-other perspectives, see Tamir and Mitchell, 2010, and Le Bouc *et al.*, 2012.)

A central idea of the ToMM-plus-selection-processing model proposed by Leslie and Polizzi (1998; see also Friedman and Leslie, 2004b; and Leslie, German and Polizzi, 2005) was that the theory-of-mind mechanism spontaneously advances two hypotheses for potential belief attribution—a 'true-belief' and a 'false-belief'—without always being able to resolve the uncertainty. In the low-demand false-belief condition, subjects calculate a (false) belief content based on informational (e.g. visual) access but do not know the current location of the object. This means they cannot attribute a 'true belief' with specific content, making its default attribution easier to overcome. In the high-demand false-belief condition, by contrast, the 'true belief' content is fully specified by their own belief, making its default attribution harder to inhibit. In this condition, subjects' anticipatory gaze is driven equally by the 'true belief' and the 'false-belief.' The result is a third pattern in looking evenly split between the two possible actions/windows reflecting two possible beliefs the agent might have. Previous studies measuring children's verbal responses could reveal only the forced outcome of selection. Spontaneous eye gaze can provide direct evidence that the theory of mind mechanism actually computes two candidates, a default 'true belief' and an access-dependent 'false belief' by catching the reasoner in a state of indecision. Findings of an early true-belief prior are compatible either with that prior being innate or with early and rapid learning that people's beliefs are typically true.

4.2 Six Measures of Processing Demands in Belief-desire Reasoning

One striking feature of the new wave of findings on early success in theory of mind is that they are derived from very different measures than those on which the old wave was based. One obvious difference is the role of language in instructing the subject, presenting the task, and eliciting responses. Whereas the standard Sally-Anne task depended upon language for each of these roles—not surprisingly performance was highly correlated with verbal ability—the Onishi and Baillargeon (2005) task did not require language at any point.

Use of language is not the only difference between traditional and 'new wave' measures. We will discuss six types of tasks for studying infants', toddlers', children's, and adults' belief-desire reasoning. The first measure is priming (Kovács *et al.*, 2010). Priming effects supported by belief metarepresentations could occur as soon as a candidate belief is entertained and before any selection process. There should be no effect of demand in priming tasks because true-belief and false-belief contents are entertained simultaneously and can prime simultaneously. This prediction is untested.

A second measure is looking time; it has been widely used with infants over many decades to address a huge variety of questions but was first applied to the false-belief task by Onishi and Baillargeon, 2005. Since then a number of looking time studies show success with high demand scenarios (e.g. He, Bolz and Baillargeon, 2011; Song, Onishi, Baillargeon and Fisher, 2008; Träuble, Marinovic and Plauen, 2010). Success with high demand, somewhat ironically, makes it difficult for looking time to test low demand, because infants should succeed with that too. Moreover, various studies suggest that looking time measures often reveal infants' competence earlier than measures that require specific action plans, such as reaching for hidden objects (Baillargeon and Graber, 1988; Diamond, 1985). Looking-time might be less demanding overall than anticipatory eye-gaze (Daum, Attig, Gunawan, Prinz and Gredebäck, 2012); perhaps this is why infants succeed even with high-demand. However, as noted earlier, there is some suggestive evidence (Scott *et al.*, 2010; Yott and Poulin-Dubois, 2012) that processing demands can nevertheless exert some influence on looking-time measures.

We have reported here on demand effects in non-verbal anticipatory eye-gaze tasks (also see Schneider, Bayliss, Becker and Dux, 2012; Schneider, Lam, Bayliss and Dux, 2012). In verbal tasks, spontaneous anticipatory eye-gaze has a longer history starting with a series of papers by Garnham (née, Clements) and colleagues. Following a verbal narrative, children's anticipatory looking was triggered by a verbal prompt, 'I wonder where Sally is going to look' (Clements and Perner, 1994; Clements, Rustin and McCallum, 2000; Garnham and Perner, 2001; Garnham and Ruffman, 2001; Ruffman, Garnham, Import and Connolly, 2001; Low, 2010; He, Bolz and Baillargeon, 2012). Three-year-olds typically succeed in these 'high-demand' tasks (while failing a direct question). The effect of varying inhibitory demand in these tasks is so far unstudied (but see Wang, Low, Jing and Qinghua, 2012, for evidence that children performed better when the processing demand was low).

A recent study by Scott *et al.* (2012; see also Barrett *et al.*, 2013) measured preferential looking at picture books while listening to a story. Children sat in front of two large storybooks placed side-by-side. One series of pictures, alternating randomly from book to book, was consistent with the narration while the other was not. For example, the narrator said, 'Emily has an apple', while one book showed a picture of a girl holding an apple and the other a picture of a girl holding a banana. Children heard a story similar to the Sally-Anne task. Finally, the narrator said, 'Emily is looking for her apple', while one picture showed Emily searching in the apple's original location (consistent with her false belief), and the other Emily searching the apple's current position. Children as young as 2.5 years preferred looking at the picture consistent with the girl's false belief rather than the real location. Again, the effect of varying inhibitory demand in these tasks awaits study.

Finally, the role of inhibitory demand on standard verbal false belief has been documented and extends into adulthood (e.g. Friedman and Leslie, 2004a, 2004b, 2005; German and Hehman, 2006; Leslie, German and Polizzi, 2005; Leslie and Polizzi, 1998; Russell, Mauthner, Sharpe and Tidswell, 1991; Surian and Leslie, 1999). In short, an array of false-belief tasks that tap core ToM competence are now available; going forward, these will be used to triangulate the developing processing systems.

4.3 Is Early Theory of Mind Real Theory of Mind?

Our findings add to the new wave that challenges twenty-five years of near consensus in the field. The old consensus held that, around four or five years, with sufficient language skills and by explicit reasoning the child constructs a theory that people have representational mental states. Furthermore, this achievement was necessary for passing the Sally-Anne task. A number of variations on this common idea added auxiliary hypotheses, executive functions, working memory, conversational skills, sibling interactions, maternal talk about the mind, or other social inputs as further necessary or sufficient factors. However, none of those views predicted that infants or toddlers would, or ever could be, capable of passing the Sally-Anne task.

As is well known, an exception to the above consensus was the ToMM theory. This proposed that a modular neurocognitive mechanism is part of core cognition and becomes functional during the second year of life. By spontaneously directing attention to the invisible, informational inner states of others and representing them, ToMM kick-starts the naïve brain into learning about mental states. Without this, such states would hardly be noticeable. By ensuring that such states are in fact noticed and attended—prior to language and without instruction—ToMM allows the young brain to learn about these states, their causes, effects, and interrelations. How does it do this? The ToMM-equipped young brain attends to (perceptible) behavior and spontaneously tries to infer the specific mental state (with content) from which the behavior issues. The key to this capacity is the "M-representation," a specialized recursive neuro-computational code required for representing the conclusions of such inferences (Leslie, 1987, 1994, 2000; see Carruthers, 2013, for an insightful defense of this view).

The ToMM theory predicted an early, spontaneous, and implicit core to theory of mind. More importantly, it made those predictions on a principled basis, namely, of a modular architecture in development. Not surprisingly, strenuous efforts are underway to revive the old consensus. The main line of argument is that early success does not reflect 'real theory of mind'. In some versions, infants are said either to possess innately or to learn associatively only low-level 'behavioral rules' tying the specific stimuli to the specific behaviors portrayed in specific experiments (Perner and Ruffman, 2005). In this vein, infants are said to react only to low-level properties of the stimuli used in specific experiments (Heyes, 2014a, 2014b). In other versions, infants lack the concept of belief and instead employ something called 'belief-like registrations' (Apperly and Butterfill, 2009). 'Real' or 'full-blown' theory of mind must wait until the child reaches four or five.

How are we to decide whether and when a child has 'real' theory of mind and acquires 'genuine' mental state concepts? One tradition in developmental work is to stipulate answers to such questions and then to see how the child measures up to our standards over time. An alternative approach is to develop and test explanatory theories of the unfolding empirical evidence and hope to discover how the processes of developmental cognition result in the adult state. In this light, judging "genuineness" of earlier states seems less compelling. In the latter vein, we offer the following reading of the currently available evidence regarding infant-toddler belief-desire reasoning.

First, a number of experiments suggest that across a range of scenarios infant-toddlers can predict an agent's action, *A*, assuming a belief, *B*, and a desire, *D*. This is essentially the basic plot of the Sally-and-Anne task and was first demonstrated by Onishi and Baillargeon (2005). Subsequently this inference was demonstrated in other scenarios (e.g. Scott *et al.*, 2010, for cup shaking; Song and Baillargeon, 2008 for misleading appearance; and Surian *et al.*, 2007, for crawling to a hidden target) and with other measures (e.g. anticipatory eye-gaze: Southgate *et al.* 2007).

Second, infants have been shown to infer an agent's desire, *D*, assuming an action, *A*, and a belief, *B* (spontaneous helping, Buttelmann *et al.*, 2009; and Buttelmann, Over, Carpenter and Tomasello, 2014). This is an especially interesting and important finding because it is a rare demonstration of an infant 'explanation task'. Rather than requiring the subject to predict an action, an explanation task presents the action and 'asks' for either a belief or a desire that would explain the action. In Buttelmann's task, a striking feature is that the action presented to the infant 'for explanation' is identical in both true-belief and false-belief conditions. What distinguishes the conditions is only the actor's belief, which the infant must infer. This maroons alternative explanations of the 'associating behaviors with stimuli' sort because the actor's behavior is identical in both conditions. Nevertheless, infants infer in the two conditions that identical behaviors reflect very different underlying intentions (desire inferred, given observed behavior and inferred belief).

Third, infants have been shown to infer the content of an agent's belief, B, assuming an action, A, and a desire, D (Southgate, Chevallier and Csibra, 2010; for related findings with toddlers and children, see Carpenter, Call and Tomasello, 2002, and Happé and Loth, 2002). In order to infer what a novel word refers to, infants have to infer which of two novel objects the speaker believes is currently in a box at which the speaker is pointing while uttering the word. In the Southgate study, an actor pointed to one box and labeled the object hidden in the box 'sefo'; meanwhile, unbeknownst to the actor, the toys in the two boxes had been switched (false-belief condition). The action is the ostensive act cum verbal label, 'sefo', and the infant assumes the actor-speaker wants the particular object she believes she is pointing at (whichever object the 'sefo' is). The infant has no prior knowledge about 'sefo', and the only way to determine the reference of 'sefo' is to infer what the actor-speaker thinks is the 'sefo', namely, the content of her belief about what she is pointing at. Southgate *et al.* (2010) found that infants solved for the value of *B*, and correspondingly their answers differed across true- and false-belief conditions.

In sum, there is evidence that infants-toddlers can carry out all three parts of a 'triad of inference':

$A \propto D, B$: infer action as a function of desire and belief
$D \propto A, B$: infer desire as a function of action and belief
$B \propto A, D$: infer belief as a function of action and desire

Subject—as always—to further experimental and theoretical investigation, we find this pattern of early abilities striking and are inclined not to dismiss it as coincidence (see also Hamlin, Ullman, Tenenbaum, Goodman and Baker (2013) on the relation between mentalistic social evaluation and inverse planning). Instead, we think it tells us something important about the developmental engine of 'genuine' theory of mind and the nature of commonsense concepts. We propose that these inferential patterns, together with the true-belief default, are hallmarks of core theory-of-mind competence.

Department of Psychology and Center for Cognitive Science
Rutgers University

References

Altmann, G. 2011: Language can mediate eye movement control within 100 milliseconds, regardless of whether there is anything to move the eyes to. *Acta Psychologica*, 137(2), 190–200.

Apperly, I.A. and Butterfill, S.A. 2009: Do humans have two systems to track beliefs and belief-like states? *Psychological Review*, 116, 953–70.

Baillargeon, R. and Graber, M. 1988: Evidence of location memory in 8-month-old infants in a nonsearch AB task. *Developmental Psychology*, 24(4), 502–11.

Baillargeon, R., Scott, R.M. and He, Z. 2010: False-belief understanding in infants. *Trends in Cognitive Sciences*, 14, 110–8

Baker, C. L., Saxe, R. and Tenenbaum, J. B. 2009: Action understanding as inverse planning. *Cognition*, 113(3), 329–49.

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

Barrett, H. C., Broesch, T., Scott, R.M., He, Z., Baillargeon, R., Wu, D., Bolz, M., Henrich, J., Setoh, P., Wang, J. and Laurence, S. 2013: Early false-belief understanding in traditional non-Western societies. *Proceedings of the Royal Society B: Biological Science*, 280(1755), 20122654.

Bartsch, K. 1996: Between desires and beliefs: young children's action predictions. *Child Development*, 67, 1671–85.

Bloom, P. and German, T.P. 2000: Two reasons to abandon the false belief task as a test of theory of mind. *Cognition*, 77, B25–B31.

Buttelmann, D., Carpenter, M. and Tomasello, M. 2009: Eighteen-month-old infants show false belief understanding in an active helping paradigm. *Cognition*, 112, 337–42.

Buttelmann, D., Over, H., Carpenter, M. and Tomasello, M. 2014: Eighteen-month-olds understand false beliefs in an unexpected-contents task. *Journal of Experimental Child Psychology*, 119, 120–6.

Carlson, S.M., Moses, L.J. and Breton, C. 2002: How specific is the relation between executive function and theory of mind? Contributions of inhibitory control and working memory. *Infant & Child Development*, 11, 73–92.

Carpenter, M., Call, J. and Tomasello, M. 2002: A new false belief test for 36-month-olds. *British Journal of Developmental Psychology*, 20(3), 393–420.

Carruthers, P. 2013: Mindreading in infancy. *Mind & Language*, 28, 141–72.

Cassidy, K.W. 1998: Three- and four-year-old children's ability to use desire- and belief-based reasoning. *Cognition*, 66, B1–B11.

Clements, W.A. and Perner, J. 1994: Implicit understanding of belief. *Cognitive Development*, 9, 377–95.

Clements, W.A., Rustin, C.L. and McCallum, S. 2000: Promoting the transition from implicit to explicit understanding: a training study of false belief. *Developmental Science*, 3, 81–92.

Cohen, A.S. and German, T. 2009: Encoding of others' beliefs without overt instruction. *Cognition*, 111, 356–63.

Daum, M. M., Attig, M., Gunawan, R., Prinz, W. and Gredebäck, G. 2012: Actions seen through babies' eyes: a dissociation between looking time and predictive gaze. *Frontiers in Psychology*, 3(370). doi: 10.3389/fpsyg.2012.00370

Dennett, D.C. 1981: Three kinds of intentional psychology. In R. Healey (ed.), *Reduction, Time and Reality*. Cambridge: Cambridge University Press, 37–61.

Diamond, A. 1985: Development of the ability to use recall to guide action, as indicated by infants' performance on AB. *Child Development*, 56, 868–83.

Fodor, J.A. 1992: A theory of the child's theory of mind. *Cognition*, 44(3), 283–96.

Friedman, O. and Leslie, A.M. 2004a: A developmental shift in processes underlying successful belief-desire reasoning. *Cognitive Science*, 28, 963–77.

Friedman, O. and Leslie, A.M. 2004b: Mechanisms of belief-desire reasoning: inhibition and bias. *Psychological Science*, 15, 547–52.

Friedman, O. and Leslie, A.M. 2005: Processing demands in belief-desire reasoning: inhibition or general difficulty? *Developmental Science*, 8, 218–25.

Gallistel, C.R. 2009: The importance of proving the null. *Psychological Review*, 116, 439–53.

Garnham, W.A. and Perner, J. 2001: Actions really do speak louder than words—but only implicitly: young children's understanding of false belief in action. *British Journal of Developmental Psychology*, 19, 413–432.

Garnham, W.A. and Ruffman, T. 2001: Doesn't see, doesn't know: is anticipatory looking really related to understanding of belief? *Developmental Science*, 4, 94–100.

German, T. and Hehman, J.A. 2006: Representational and executive selection resources in 'theory of mind': evidence from compromised belief-desire reasoning in old age. *Cognition*, 101, 129–52.

Gopnik, A. and Wellman, H.M. 1995: Why the child's theory of mind really is a theory. In M. Davies and T. Stone (eds), *Folk Psychology: The Theory of Mind Debate*. Oxford: Blackwell, 232–58.

Hamlin, J.K., Ullman, T., Tenenbaum, J., Goodman, N. and Baker, C. 2013: The mentalistic basis of core social cognition: experiments in preverbal infants and a computational model. *Developmental Science*, 16, 209–26.

Happé, F. and Loth, E. 2002: 'Theory of mind' and tracking speakers' intentions. *Mind & Language*, 17, 24–36.

He, Z., Bolz, M. and Baillargeon, R. 2011: False-belief understanding in 2.5-year-olds: evidence from change-of-location and unexpected-contents violation-of-expectation tasks. *Developmental Science*, 14, 292–305.

He, Z., Bolz, M. and Baillargeon, R. 2012: 2.5-year-olds succeed at a verbal anticipatory-looking false-belief task. *British Journal of Developmental Psychology*, 30, 14–29.

Heyes, C. 2014a: False belief in infancy: a fresh look. *Developmental Science*, 17, 647–59.

Heyes, C. 2014b: Submentalizing: I am not really reading your mind. *Perspectives on Psychological Science*, 9, 131–43.

Kikuno, H., Mitchell, P. and Ziegler, F. 2007: How do young children process beliefs about beliefs? Evidence from response latency. *Mind & Language*, 22, 297–316.

Kovács, A.M., Téglás, E. and Endress, A.D. 2010: The social sense: susceptibility to others' beliefs in human infants and adults. *Science*, 330, 1830–4.

LeBouc, R., Lenfant, P., Delbeuck, X., Ravasi, L., Lebert, F., Semah, F. and Pasquier, F. 2012: My belief or yours? Differential theory of mind deficits in frontotemporal dementia and Alzheimer's disease. *Brain*, 135, 3026–38.

Leslie, A.M. 1987: Pretense and representation: the origins of 'theory of mind'. *Psychological Review,* 94, 412–26.

Leslie, A.M. 1992: Pretense, autism, and the 'theory of mind' module. *Current Directions in Psychological Science*, 1, 18–21.

Leslie, A.M. 1994: *Pretending* and *believing*: Issues in the theory of ToMM. *Cognition*, 50, 211–238.

Leslie, A.M. 2000: 'Theory of mind' as a mechanism of selective attention. In M. Gazzaniga (ed.), *The New Cognitive Neurosciences,* 2nd edn. Cambridge, MA: MIT Press, 1235–47.

Leslie, A.M., Friedman, O. and German, T. 2004: Core mechanisms in 'theory of mind'. *Trends in Cognitive Sciences*, 8, 528–33.

Leslie, A.M., German, T. and Polizzi, P. 2005: Belief-desire reasoning as a process of selection. *Cognitive Psychology*, 50, 45–85.

Leslie, A.M. and Polizzi, P. 1998: Inhibitory processing in the false belief task: two conjectures. *Developmental Science*, 1, 247–54.

Leslie, A.M. and Thaiss, L. 1992: Domain specificity in conceptual development: neuropsychological evidence from autism. *Cognition*, 43, 225–51.

Low, J. 2010: Preschoolers' implicit and explicit false-belief understanding: relations with complex syntactical mastery. *Child Development*, 81, 597–615.

Low, J. and Watts, J. 2013: Attributing false beliefs about object identity reveals a signature blind spot in humans' efficient mind-reading system. *Psychological Science*, 24, 305–11.

Luo, Y. 2011: Do 10-month-old infants understand others' false beliefs? *Cognition*, 121, 289–98.

Mitchell, P. 1994: Realism and early conception of mind: a synthesis of phylogenetic and ontogenetic issues. In C. Lewis and P. Mitchell (eds), *Children's Early Understanding of Mind*. New York: Cambridge University Press, 141–72.

O'Grady, C., Kliesch, C., Smith, K. and Scott-Phillips, T.C. 2015: The ease and extent of recursive mindreading, across implicit and explicit tasks. *Evolution and Human Behavior,* 36(4), 313–22.

Onishi, K. and Baillargeon, R. 2005: Do 15-month-old infants understand false beliefs? *Science*, 308, 255–8.

Perner, J. 1991: *Understanding the Representational Mind*. Cambridge, MA: MIT Press.

Perner, J. 1995: The many faces of belief: reflections on Fodor's and the child's theory of mind. *Cognition*, 57, 241–69.

Perner, J. and Ruffman, T. 2005: Infants' insight into the mind: how deep? *Science*, 308, 214–6.

Roth, D. and Leslie, A.M. 1991: The recognition of attitude conveyed by utterance: a study of preschool and autistic children. *British Journal of Developmental Psychology*, 9, 315–30.

Rubio-Fernández, P. 2013: Perspective tracking in progress: do not disturb. *Cognition*, 129, 264–72.

Ruffman, T., Garnham, W., Import, A. and Connolly, D. 2001: Does eye gaze indicate implicit knowledge of false belief? Charting transitions in knowledge. *Journal of Experimental Child Psychology*, 80, 201–24.

Russell, J., Mauthner, N., Sharpe, S. and Tidswell, T. 1991: The 'windows task' as a measure of strategic deception in preschoolers and autistic subjects. *British Journal of Developmental Psychology*, 9, 331–49.

Schneider, D., Bayliss, A.P., Becker, S.I. and Dux, P.E. 2012: Eye movements reveal sustained implicit processing of others' mental states. *Journal of Experimental Psychology: General*, 141, 433–8.

Schneider, D., Lam, R., Bayliss, A.P. and Dux, P.E. 2012: Cognitive load disrupts implicit theory-of-mind processing. *Psychological Science*, 23, 842–7.

Scholl, B.J. and Leslie, A.M. 2001: Minds, modules, and meta-analysis. *Child Development*, 72, 696–701. (Reprinted in M. Mason (ed.), *Taking Sides: Clashing Views on Controversial Issues in Cognitive Science*. New York: McGraw-Hill, 34–43.)

Scott, R.M., Baillargeon, R., Song, H. and Leslie, A.M. 2010: Attributing false beliefs about non-obvious properties at 18 months. *Cognitive Psychology*, 61, 366–95.

Scott, R.M., He, Z., Baillargeon, R. and Cummins, D. 2012: False-belief understanding in 2.5-month-olds: evidence from two novel verbal spontaneous-response tasks. *Developmental Science*, 15, 181–93.

Senju, A., Southgate, V., White, S. and Frith, U. 2009: Mindblind eyes: an absence of spontaneous theory of mind in Asperger syndrome. *Science*, 325(5942), 883–5.

Song, H. and Baillargeon, R. 2008: Infants' reasoning about others' false perceptions. *Developmental Psychology*, 44, 1789–95.

Song, H., Onishi, K, Baillargeon, R. and Fisher, C. 2008: Can an agent's false belief be corrected by an appropriate communication? Psychological reasoning in 18-month-old infants. *Cognition,* 109(3), 295–315.

Southgate, V., Chevallier, C. and Csibra, G. 2010: Seventeen-month-olds appeal to false beliefs to interpret others' referential communication. *Developmental Science*, 13, 907–12.

Southgate, V., Senju, A. and Csibra, G. 2007: Action anticipation through attribution of false belief by 2-year-olds. *Psychological Science*, 18, 587–92.

Surian, L., Caldi, S. and Sperber, D. 2007: Attribution of beliefs in 13-month-old infants. *Psychological Science*, 18, 580–6.

Surian, L. and Leslie, A.M. 1999: Competence and performance in false belief understanding: a comparison of autistic and normal 3-year-old children. *British Journal of Developmental Psychology*, 17, 141–55.

Tamir, D.I. and Mitchell, J.P. 2010: Neural correlates of anchoring-and-adjustment during mentalizing. *Proceedings of the National Academy of Sciences*, 107(24), 10827–32.

Träuble, B., Marinovic, V. and Plauen, S. 2010: Early theory of mind competencies: do infants understand others' beliefs? *Infancy*, 15, 434–44.

Wang, B., Low, J., Jing, Z. and Qinghua, Q. 2012: Chinese preschoolers' implicit and explicit false-belief understanding. *British Journal of Developmental Psychology*, 30(Pt 1), 123–40.

Wellman, H.M., Cross, D. and Watson, J. 2001: Meta-analysis of theory of mind development: the truth about false-belief. *Child Development*, 72, 655–84.

Yott, J. and Poulin-Dubois, D. 2012: Breaking the rules: do infants have a true understanding of false belief? *British Journal of Developmental Psychology*, 30(Pt 1), 156–71.

Zaitchik, D. 1991: Is only seeing really believing?: Sources of true belief in the false belief task. *Cognitive Development*, 6, 91–103.