PRESCHOOLERS’ AND ADULTS’ BELIEF REASONING AND TASK DEMAND

By

LU WANG

A thesis submitted to the

Graduate School – New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Master of Science

Graduate Program in Psychology

written under the direction of

Dr. Alan M. Leslie

and approved by

_______________________________

_________________________

_________________________

New Brunswick, New Jersey

May, 2011
Thirty-years research seemed to reveal that there is a U-shape development in children’s theory-of-mind abilities: infants have the competence to attribute false beliefs properly when measured by looking time and anticipatory eye gaze, while children younger than four systematically fail the standard false belief tasks measuring their voluntary responses. Why is it, and why does the infants’ implicit belief reasoning seem to be free from the inhibition and selection requirements? Are there really two systems, one explicit measured by verbal tasks and one implicit shown in non-verbal context, involved in theory-of-mind reasoning? We investigated preschoolers’ (experiment 1) and adults’ (experiment 2) belief attribution with the anticipatory looking method, in both a low-demand and high-demand false-belief reasoning tasks. Subjects’ nonverbal performances turned out to be better in the low-demand condition: they showed strong bias to look at the locations congruent with the actor’s false belief. However, neither adults’ nor children’s eye gaze in the high-demand condition showed belief-congruent anticipations. Besides, adults’ predictions were better with verbal responses than with nonverbal anticipatory eye gaze. Together, these results suggested that there are dual systems involved in belief reasoning, and the two systems share similar property in terms of their sensitivity to the task demands.
Acknowledgements

I would like to thank Dr. Alan Leslie, Anton Shcherbakov, Melissa Kibbe, and all members of Cognitive Development Lab, for their assistance and support in the research of my thesis. Special thanks to Dr. Victoria Southgate for her generous sharing of the original video stimuli. I also want to thank the undergraduate students, children and their parents for participation.
Table of Contents

Abstract ........................................................................................................................................... ii

Acknowledgments .................................................................................................................. iii

List of Figures ......................................................................................................................... vi

Introduction ............................................................................................................................ 1

• Overview ............................................................................................................................. 1

• 1. Three-year-olds’ failures: the story of conceptual deficit ........................................ 3

• 2. Think twice: meta-analysis and early competence ...................................................... 4

• 3. Infants’ success: the story of early competence ............................................................. 6

• 4. From early success to later failures: the theoretical explanations ............................... 9

• 5. Processing demands and the dual systems: the present study ..................................... 11

Experiment 1 .......................................................................................................................... 15

• Methods .............................................................................................................................. 15

• Results ............................................................................................................................... 18

• Discussion ......................................................................................................................... 21
Experiment 2 ................................................................. 23

• Methods ........................................................................... 23

• Results ............................................................................. 24

• Discussion ....................................................................... 29

General Discussion ............................................................. 30

• 1. ToMM + selection process and the properties of the dual systems ........ 31

• 2. VOE vs. AL discrepancy .................................................... 33

• 3. Limitations and remaining questions ........................................ 35

Conclusion ........................................................................... 36

References ........................................................................... 37

Footnotes ............................................................................. 41
List of Figures

1. Experiment 1: Key events in familiarization trials ........................................ 17
2. Experiment 1: Key events in test trials (from Senju, et al., 2009) .................. 17
3. Experiment 1: Mean of children’s DLS ....................................................... 20
4. Experiment 1: Percentage of children’s first look to different windows .......... 21
5. Experiment 2: Mean of adult’s DLS .............................................................. 26
6. Experiment 2: Percent of adults’ DLS favoring different sides .................... 27
7. Experiment 2: Percent of adults’ nonverbal & verbal responses ................. 29
Introduction

Overview

We regularly interpret people’s behavior based on their mental states. Thus, we can easily understand Simba’s fear and guilty when his father, Mufasa, was killed, and he was led to believe by Scar that it was his fault. This ability is so fundamental that developmental psychologists have long been interested in what’s the origin of it and how it develops. Typically, they use false belief understanding as evidence of mental states reasoning. (Premack & Woodruff, 1978).

The standard task to test children’s false belief understanding is the Sally-Anne task (Baron-Cohen, Leslie & Frith, 1985; Wimmer & Perner, 1983):

Sally has a marble, but she doesn’t want to play with it. So Sally puts her marble into a basket and leaves. On Sally’s absence, Anne removes her marble from the basket and puts that into another box. Where will Sally look for her marble when she comes back?

A long-standing finding is that children younger than four years of age could not reason about the false belief and say Sally would look for her marble in the box (for a review and meta-analysis, see Wellman, Cross, & Watson, 2001). However, a break-through in this field showed infants’ competence in false belief reasoning when measured by various non-verbal methodologies (Onishi & Baillargeon, 2005; Southgate, Senju, & Csibra, 2007).
In light of the infants’ competence and preschoolers’ failures, a dual-system architecture, following Leslie & Thaiss (1992), Leslie, German & Polizzi (2005), has been used to explain the belief reasoning mechanism (Baillargeon, Scott, & He, 2010): one is implicit, functioning spontaneously and present early in life, and it gives rise to infants’ competence in false belief attribution measured by looking time or anticipatory looking; the other is explicit, requiring effective executive function to avoid interference from the child’s own knowledge. This model suggests that limited inhibition resources lead to younger children’s difficulties in verbal belief tasks.

In this thesis, the role of executive processes in both children’s and adults’ implicit belief reasoning was investigated. We examine a key assumption of the above model, namely, that early success on non-verbal false belief tasks is due to the fact that these versions of the task make no executive demands and thus can be passed very early. We show that, on the contrary, the “early success” tasks do make executive demands similar to “late success” false belief tasks. Even with sufficient executive resources, adults’ implicit reasoning was not as good as their explicit performances. Together, these results suggested that executive processes are required in both explicit and implicit belief reasoning systems. Therefore, the present studies provide a more complete picture underlying false belief reasoning.

I will firstly summarize the traditional findings of children’s abilities to understand false belief, and the new field of infant’s competence in theory of mind reasoning. I’ll then outline different theories that try to account for the developmental trajectory from infancy to preschool years. Armed with that background, I shall describe the present study investigating the mechanism underlying theory of mind development.
1. Three-year-olds’ failures: the story of conceptual deficit

For a long time, it was believed that the differences of false belief reasoning between three-year-olds and four-year-olds are due to a major conceptual change in the preschool years, changing from non-representational mental states-understanding in younger children to adult-like mentalistic understanding of people in four-year-olds. For example, children younger than four years old not only failed to understand the character’s false belief of a toy’s location (like in Sally-Anne task), they also failed to appreciate their own false belief after knowing the truth (Gopnik & Astington, 1988).

However, there is an alternative explanation for younger children’s false belief errors, this is, performance demands. Young children have the competence to understand mental states, but they fail to demonstrate their understanding in standard tasks because of the linguistic complexity and misinterpretations of the verbal questions (Mitchell & Lacobche, 1991; Siegal & Beattie, 1991), or the difficulty to inhibit their representation of reality (Carlson, Moses, & Hix, 1998; Zaitchik, 1991). This claim of early competence is supported by several studies revealing three-year-olds’ success in false belief understanding. For example, Siegal & Beattie (1991) suggested that three-year-olds typically failed the change-of-location false belief tasks because they may have interpreted the question “where will Sally look for her marble?” to mean “where will Sally have to look to find her marble?” To rule out the linguistic misunderstanding, they made the question clearly about Sally’s false belief, by asking “where will Sally look first?” They found that three-year-olds performed much better to predict Sally’s initial behavior with her false belief (Siegal & Beattie, 1991). Another study used a story similar to the Sally-Anne task, but children were told the toy’s real location, instead of seeing it.
The researcher found that when three-year-olds did not see the toy’s real location, they were more likely to predict the character’s behavior based on false belief (Zaitchik, 1991; see also Wellman & Bartsch, 1988). Therefore, these studies suggested that younger children have the concept of other’s belief, but their competence is masked by several performance demands of demonstrations in the traditional tasks.

Nevertheless, the early competence was dismissed by a meta-analysis done by Wellman et al. (Wellman, Cross, & Watson, 2001). By combining hundreds of studies across different cultures and over more than a decade of time, they concluded that it was a universal principle that children’s understanding of mind underwent a three-step development, from desire psychology, desire-belief psychology, to belief-desire psychology. Specifically, two-year-olds could understand people’s desire, intentions etc. by observing people’s interactions with objects. But understanding of mind at this stage is purely based on actions ABOUT objects. Later on, three-year-olds start to realize that people also have different attitudes towards the objects, which is independent from their actions, but at the second stage of development, their understanding of the mental attitudes are not mature. Finally, benefitting from massive experiences with people, four-year-olds started to understand the mental world in an adult-like way, and they could appreciate people’s thoughts independently from their behaviors. They suggested that the conceptual change from desire psychology to belief-desire psychology underlay the observed developmental trajectory of preschoolers’ false belief reasoning (Flavell, 1999; Wellman, et al, 2001).

2. Think twice: meta-analysis and early competence
The universal time table was non-controversially consistent with the conceptual change theory. However, were the results from meta-analysis able to exclude the possibility of early competence, such as suggested by the theory of mind mechanism and selection process theory? The answer is no.

According to theory of mind mechanism and selection process theory (ToMM/SP), the ability to understand other’s mental states has a specific innate basis (Leslie, 1987, 1994; Leslie & Thaiss, 1992), probably as a modular system (ToMM). Similar to other modules (Fodor, 1983), ToMM is specifically sensitive to certain inputs and calculates specialized outputs, the representation of mental states. However, the innate basis is quite limited and not-mature, and it’s open to development given the supplements from general processing abilities, most importantly, the inhibitory selection abilities (Selection Process, SP) (Leslie, Friedman, & German, 2004; Scholl & Leslie, 2001). In the standard false belief task, children need to inhibit the pre-potent representation of the truth in favor of reasoning about other’s false belief, thus, an effective inhibitory selection process (SP) besides the representation of belief is necessary for success. Given that the development of inhibition abilities is slow in childhood (Diamond, Kirkham, Amso, 2002) and parallels the observed development of theory-of-mind performances from three-to-four years of age, the results of the meta-analysis are compatible with ToMM/SP theory.

Secondly, the meta-analysis also suggested that improved performances in modified false belief tasks were not strong enough to claim children’s early competence, since only sparse modifications were able to make an effect, and even when there were improvements, young children’s performances were far from perfect (Wellman, et al., 2001). However, early competence theory does not guarantee younger children’s success.
or above-chance performances in the standard false belief task. As Bloom & German (2000) pointed out, passing the standard false-belief task requires a lot more than theory-of-mind abilities. It’s sufficient to claim one’s theory-of-mind ability if he/she could pass the false belief tasks, but it’s not necessary so (Scholl & Leslie, 2001). For example, it could be that the general structure of the tasks measuring children’s verbal responses caused the limitations.

In sum, children’s failures and the universal developmental pattern from three-to-four years of age does not close the door for early competence. The contrast between the conceptual change and the early competence theory may not be about development, but the origins of theory-of-mind abilities. Consequently, only the early competence theory predicts that younger children in their second year of life (Leslie, 1994) will be able to reason about false belief with sensitive measurements and tasks requiring lower processing demands. Since the high processing demands may be inevitable in the verbal tasks and two-year-olds do not have the abilities to understand the stories, it would be more informative to investigate younger children’s belief reasoning with non-verbal measurements.

3. Infants’ success: the story of early competence

Eye gaze and preferential looking time have been used substantially as reliable non-verbal measurements to look at younger children’s cognitive abilities. A lot of studies measuring infants’ eye gaze were able to reveal infant’s reasoning about physical rules (e.g. Baillargeon, 1987). Clements and Perner (1994) were the first to apply the implicit anticipatory eye gaze measurement to investigate children’s false belief reasoning. They
used a change-of-location false belief task: a mouse put his cheese in one blue box and went to bed; but on his absence, another mouse came in and moved his cheese into the red box. They asked two-to-four-year-olds to predict which box the first mouse would open when he came back. Besides, they added an anticipation prompt “I wonder where he is going to look” before the mouse reappeared via the mouse holes (two mouse holes, leading to the two boxes respectively), and they left one to two seconds after the prompt to measure the children’s anticipation by eye gaze. Children who were nearly three years old shifted their eye gaze to the mouse hole they expect the mouse to appear, and their anticipations showed correct attribution of false belief, even though they failed by the explicit responses immediately after the prompt (either by words or pointing) (Clements & Perner, 1994).

But the nonverbal measurements in Clements and Perner’s (1994) task was embedded in a verbal context. The strongest evidences for early competence came from studies with complete non-verbal materials and measurements, exploring infants’ false belief reasoning. For example, Onishi and Baillargeon (2005) adapted the Sally-Anne task into a nonverbal violation-of-expectation task (VOE) and showed it to 15-month-olds. In the beginning of the experiment, an actor hid a toy watermelon in one box and reached for it later (a green box and a yellow box in total). In the test trial, the actor was absent when the toy moved from one box to the other, therefore, the actor had a false belief about its location. In the end, the actor either searched for the toy in the box where she falsely believed it was or the box where the toy really was. They found that 15-month-olds looked reliably longer when the actor with a false belief reached for the toy’s real location. Subsequent VOE studies confirmed and extended this finding, and these studies
suggested that infants in the second year of life were already able to attribute to an agent a false belief of a toy’s location (e.g. Trauble, Marinovic, & Pauen, 2010), an object’s perceptual appearance or identity (Scott & Baillargeon, 2009; Song & Baillargeon, 2008), and even an object’s non-obvious property (Scott, Baillargeon, Song, & Leslie, 2010).

Most recently, with the new technology of flexible eye tracking with children, Southgate et al. (2007) showed several nonverbal videos to 25-month-olds and assessed infants’ anticipatory eye gaze with an eye-tracking monitor. Their task was a nonverbal adaption built on Clements and Perner’s (1994) work. In the videos, an agent sat at the back of a panel with her head visible to the subjects. There were two boxes and a closed window above each box. At the beginning of the familiarization videos, a puppet showed up with a toy, and hid the toy in one of the boxes when the agent was watching. Then the window was illuminated, and the agent opened and reached through the correct window to get the toy. In the following test trial, the puppet hid the toy in the Left box, then he paused for a second, retrieved the toy from the Left box and placed it into the Right one while the agent was watching. Then, the agent was distracted by a phone ring and turned away from the display. When she was facing away, the puppet showed up again and removed the toy from the Right box and took it completely out of the scene. The phone ring stopped right after the puppet disappeared, and the agent turned back immediately. Then the windows lit up, signaling that the agent was going to search for the toy. According to the recording from the Tobii eye tracker, seventeen out of twenty 25-month-olds shifted their eye gaze to the Right window immediately after the signal. They made the anticipation of the agent’s behavior by their first eye gaze which could only be explained by a false belief attribution to the agent. The anticipatory eye gaze strongly suggested that
children as young as two years old had the competence to understand other’s false belief and make efficacious behavioral predictions based on that (Southgate, Senju, & Csibra, 2007).

4. From early success to later failures: the theoretical explanations

Thus, there seems to be a U-shape development from infants to four-and-five-year-olds. Specifically, infants in their second year of life are ready to understand people’s mental worlds: they look longer if a person’s behavior is not compatible with their belief status, and they can correctly anticipate people’s actions by attributing an appropriate belief to them. Nevertheless, three-year-olds typically failed to reason about other’s false belief, and they usually followed their own knowledge of the truth when asked to predict other’s behavior. Finally, four-year-olds and older children started to reason about other’s false belief in an adult-like way, and they were able to pass the standard verbal tasks constantly. Then, why do three-year-olds struggle and fail the tasks they are supposed to be competent at as younger infants? Apparently, the idea of a conceptual deficit in younger children cannot account for the developmental trajectory, given infants’ performances.

A dual-systems account, invoking the ToMM/SP theory, was used to explain the U-shape development (Baillargeon, Scott, & He, 2010; Leslie, Friedman, & German, 2004; Leslie, German, & Polizzi, 2005; Leslie & Polizzi, 1998; Scott, Baillargeon, Song, & Leslie, 2010). According to this account, the non-verbal tasks revealing early success and the verbal tasks showing three-year-olds’ failures tapped into two distinct cognitive processes: an early-developed, intuitive, and probably unconscious one driven by a
modularized mechanism (ToMM, theory-of-mind mechanism), and a late-developed, voluntary, and explicit one facilitated by general inhibitory ability and response-selection process (SP, selection process). On the one hand, there is an innate specialized module, ToMM, enabling infants to attribute context-appropriate beliefs to others. On the other hand, the standard false-belief tasks require elicited responses from children and additional processes beyond the implicit false belief representation (such as response-selection and prepotent response-inhibition), hence impose a high processing demand which younger children do not overcome until they are four years old.

The dual-systems theory could explain the available evidence from infants to preschoolers, and even neuro-atypical development in autism. There is a good amount of evidence showing that children with autism have specialized problems in understanding theory of mind. For example, autistic children failed the standard false belief task, while the control group with similar mental age, including normal ones and children with Down’s syndrome, were able to pass (Baron-Cohen, Leslie, & Frith, 1985). Besides, modifications found to enhance three-year-olds’ performance, such as the “look first” question (Siegal & Beattie, 1991) described above, had little effect on autistic children’s performances, suggesting a different source of problems underlying autistic children’s failures in the standard task (Surian & Leslie, 1999). What’s more, the nonverbal studies provided the most straightforward evidence for the specialized impairments of the domain-specific module in autism. Senju et al. (Senju, Southgate, White, & Frith, 2009) used the same anticipatory task as used in Southgate et al. (2007), and showed the nonverbal videos via the Tobii eye tracker to a group of adults with Asperger syndrome. The task was able to measure infants’ spontaneous eye movement to predict the actor’s
behavior based on her false belief. However, adults with Asperger syndrome failed to demonstrate such spontaneous anticipations, yet they were able to reason about false belief explicitly in standard verbal tasks. Similar results were found with autistic children (Senju, et al., 2010).

Therefore, children with autism spectrum disorder and adults with Asperger syndrome may be able to pass the standard false belief tasks, but they fail to do so spontaneously with their eyes. By contrast, younger children may fail to reason about false belief explicitly in the standard tasks, but they are able to do so implicitly from the second year of life measured by non-verbal tasks. The double dissociations lent further support to the dual-systems account.

5. Processing demands and the dual systems: the present study

What happens when children try to apply their implicit competence voluntarily? What is responsible for the gap between the two systems?

It is implied in the dual-systems theory that the processing required by the standard tasks makes them difficult for three-year-olds. In the standard tasks, such as the Sally-Anne task, the presence of the target and children’s knowledge of the reality would be very salient, and it serves as the potent response. Children will have to resist making inferences from the potent contents to correctly reason about Sally’s false belief, and it’s hard for three-year-olds who have limited processing resources. When not elicited to infer Sally’s belief, as investigated by non-verbal measurements, the theory-of-mind module is effective enough to calculate the correct content of Sally’s belief. In other words, the implicit belief attribution requires minimal processing resources, such that
without effective inhibition or response selection abilities, even infants are able to reason about Sally’s belief correctly.

However, the representation of the truth is present in the non-verbal situations. So, it’s mysterious why children’s knowledge of the truth seems not to be salient, and fails to bias infants reasoning in the non-verbal tasks.

One possibility is that the implicit reasoning system powered by the theory-of-mind mechanism provides completely demand-free processing. The specialized module, by its nature, will calculate the appropriate belief contents. Without any elicited requests, infants and younger children would simply follow the calculation outcomes of the mechanism, and attribute the correct belief to others. However, when asked about other’s belief, both the false belief representation and the representation of the world are available for use. Given the salience of the truth and the younger children’s limited processing resources, they fail to select the correct false belief content. According to this view, the two systems are completely different in terms of their vulnerability to the representations of truth.

Another possibility is that some processing resources are still needed for the implicit system to correctly attribute belief contents to others. The differences observed between verbal failures and nonverbal competences might be caused by the structure of the tasks. The standard verbal tasks require a lot of processing resources, from correctly interpreting the questions to generating proper explicit responses, leaving few resources for three-year-olds to inhibit the potent representation of the reality. The nonverbal tasks, however, free up infants’ and younger children’s limited processing resources, which
may be sufficient to reduce the salience of reality. According to this view, the implicit and explicit systems are quite similar. In fact they are uniform, both controlled by the same mechanisms, namely the highly specialized ToMM along with the less specialized selection process to apply their processing resources (Leslie, German, & Polizzi, 2005).

Apparently, both views are successful in explaining the available evidence. To better understand the cognitive mechanisms underlying the core ability of theory-of-mind, we need to know more about the property of the non-verbal belief reasoning system. Specifically, will the implicit reasoning system be subjected to the processing demands required by tasks?

To investigate this question, we need to find out the relative processing demands of different tasks, and what variations determined the demands. Fortunately, decades of research in children’s performances in the standard tasks provided a rich amount of information for task analysis. As summarized in part one, there were several modifications of the standard tasks reported to enhance younger children’s performances, by reducing the processing demands, such as resolving the linguistic confusion in the questions by asking “where to look first” (Siegal & Beattie, 1991), or eliminating the potency of reality by telling but not showing children the target’s real place (Zaitchik, 1991). However, these modifications were based on varying children’s verbal inputs, and they were not ready to be adapted to a nonverbal task. Another effective modification is the non-presence of the desired target. For example, instead of having the desired target hidden on the scene, if it was out of the scene completely, but the actor falsely believed it in one of the present boxes, three-year-olds were more able to predict the actor’s behavior based on her false belief (Bartsch, 1996). The effect of a non-present target on children’s
enhanced performances of belief reasoning was also confirmed in the meta-analysis (Wellman, et al., 2001).

Notice that the non-present modification could be well applied to a non-verbal context. In fact, this modification has been used in a nonverbal study: in Southgate et al. (2007), a puppet replaced a toy’s location when the actor was facing away, and eventually, the puppet took the toy completely out of the scene. The authors showed the events as nonverbal videos, and used an eye tracker to measure children’s anticipatory eye gaze. Inspired by the corresponding verbal studies, we could simply increase the processing demand of the task by leaving the toy on the scene: instead of having the puppet remove the toy completely, he would replace the toy and leave it hidden in the other location when the actor is facing away. The presence of the target imposes higher processing demands, corresponding to the more difficulties of the standard task (e.g. Sally-Anne task) than the non-presentation modification. Then, the question is, will subjects’ performances be sensitive to the increased processing demands of the nonverbal task, measured by their spontaneous eye gaze?

This is the aim of the present study. We used the Tobii eye tracker to show the nonverbal videos to subjects. The processing demands of the videos are manipulated by the presence of the target: the target is either removed completely (low-demand, as used in Southgate et al., 2007) or hidden in one of the boxes on the scene (high-demand). If the implicit reasoning system is processing-free, then we would not expect any differences between the low-demand and high-demand conditions. Subjects should correctly anticipate the actor’s behavior based on her false belief in both conditions. However, if the implicit system measured by nonverbal tasks is also sensitive to the processing
demands of the tasks, then we would expect subjects to be better at false belief anticipation in the low demand condition. Either way, the present study will shed light on the property of the implicit theory of mind system[1].

Experiment 1

Method

70 two-to-three-year-olds participated in the study (mean age was 36.5 months, range: 24.2 – 47.9 months, 38 females). They were randomly assigned to one of the three conditions: 26 in False-Belief Low-demand (LD), 26 in False-Belief High-demand (HD), and 18 in True-Belief condition (TB). Another 16 subjects were excluded due to failure to meet gaze criteria: specifically, to ensure that the subjects encoded sufficient information and made eye gaze towards either window, we calculated their total looking time to both windows that may indicate their belief attribution. Less looking time may indicate the loss of tracking from the Tobii eye tracker, head movements of the subjects, or subjects’ poor attention to the scenes, therefore, it may not reflect subjects anticipation by eye gaze accurately. The total looking time was required to be longer than 150 milliseconds to be included.

We used a Tobii (Stockholm, Sweden) T60 XL Eye Tracker to present the stimuli and record the subjects’ eye gaze in the tasks. The study and gaze recording were controlled by a computer with Tobii Studio presentation software. Children either came with their

[1]: The implicit system is usually measured by non-verbal tasks and the explicit reasoning is measured in verbal tasks, therefore, implicit/explicit and non-verbal/verbal were used interchangeable here.
parents to the lab or participated in the study at their preschool. In the lab, children sat on their parent’s lap, 60 to 80 cm away from the monitor (adjusted in terms of the tracking quality in calibration), in a separate booth surrounded by a curtain. A 5-point moving toy calibration was applied before the experiment.

Each subject watched videos showing two main familiarization trials in which a puppet hid a toy in one of two boxes in the scene, and after a chime/illumination signal, an actor reached through the window behind the box to retrieve the toy. In the following test trial, while the actor had her back turned, the puppet either transferred the toy to the other box (High Demand, HD) or retrieved the toy and left with it (Low Demand, LD). The actor then turned back and the signal indicated that she was about to search. In the True Belief (TB) condition, the actor turned back when the puppet was still moving the toy, and she saw that the toy was removed and put into the other box (see Fig.1 for familiarization trial and Fig.2 for test trial).

The videos used in the LD condition were the same as those used by Southgate (Southgate, et. al., 2007), and the HD videos were obtained by editing the LD ones such that instead of leaving with the toy, the puppet removed the toy and put that into the other box. True-Belief videos were also obtained by editing in which the actor turned back early and caught the puppet when it was still moving the toy.
Fig. 1 Key events in familiarization trials. All subjects watched the Left-box familiarization trial first, then the Right-box familiarization trial (Left-box familiarization shown in the figure).

Fig. 2 Key events in test trials (from Senju, Southgate, White, & Frith, 2009). In the False Belief condition, the actor was distracted and facing away when the puppet removed the toy. In the High-Demand (HD) task, the puppet put the toy into the other box and left; in the Low-Demand (LD) task, the puppet left with the toy. In True Belief (TB) task, the actor turned back when the puppet was moving the toy, and she saw that the toy was removed and put into the other box.

The actor’s belief of the toy’s location was counterbalanced across subjects (side). During the whole process when the actor was facing the display, she followed the
puppet’s movement very tightly and clearly with obvious head’s movement. At the end of each video, the windows lit up accompanied by a chime. Then the video paused for four seconds with the windows illuminated, while the eye tracker recorded subjects’ first saccade to the windows and measured how long they looked at either window.

**Results**

Subjects’ nonverbal performances were measured using the Differential Looking Score (DLS), a measure of their preference to look at the window congruent with the actor’s false belief over the four seconds. DLS was calculated by subtracting the time subjects spent looking at the false-belief window from that to the true-belief window (that is, where the actor would search if she knew the truth in both HD and TB conditions, and the non-false-belief window in the LD condition), and then dividing it by the sum of the looking time to both sides. Therefore, DLS would vary from -1 to 1: a DLS close to 0 meant the subjects spent equal amount of time looking at true-belief and false-belief windows, and a DLS above 0 indicated that subjects were more likely to look at the false-belief window within the four seconds of anticipation.

Preliminary analysis revealed that children did slightly better in the Right-side event in which the actor falsely believes that the toy is hidden in the Right side box. But the effect of side did not reach significance. Gender did not play a role in subjects’ eye gaze performances either. Therefore, we pooled over these factors in the following analysis.

The familiarization trial prior to the test was used to establish the contingency between the illumination of the windows and the actor’s subsequent searching behaviors. Previous studies used subjects’ eye gaze in the last familiarization trial as a screening criterion,
such that only those subjects who made their first saccade towards the correct window (first look in familiarization, Fam-FL) or looked longer at the correct window (preferential look in familiarization, Fam-PL), showing that they understood this contingency, were included (Southgate et al., 2007; Senju et al., 2010). However, neither of the criteria in the familiarization trial had an effect on subjects’ nonverbal belief reasoning in the present study (according to the 2 by 3 ANOVA with each familiarization criterion respectively and the demand of tasks as two independent variables, $p = .666$ for Fam-FL, $p = .197$ for Fam-PL). Therefore, we did not use it to exclude subjects.

2 (side) by 3 (task demand) ANOVA on children’s DLS showed that only the main effect of demand was significant, $F (2, 64) = 4.89, p < .05, \eta^2 = .133$. In the LD condition they preferred to look longer at false-belief window (mean DLS = 0.19, $Z = 1.69, p < .05$, one-tailed) and showed ambiguous preferences in the HD condition (DLS = -0.11). The difference between LD and HD condition was significant, $t (50) = 1.76, p < .05$ (one-tailed). In the TB condition, there were more likely to look at the true-belief window (mean DLS = -0.34, $t (17) = -2.56, p < .05$), and it was significantly different from their preference in LD, $t (42) = 2.95, p < .01$. However, children’s ambiguous preference in the HD condition was not different from their preferential looking in TB (see Fig.3).
Fig. 3 Mean of children’s Differential Looking Scores. They looked significantly longer at the belief-congruent window in LD and TB conditions over the last four seconds of anticipation, but their preference was ambiguous in HD condition.

*: $p < .05$ (one-tailed); **: $p < .05$.

In terms of subjects first gaze shift towards either window, 13 out of 18 subjects looked at the true-belief window first in the TB condition, significantly above chance, $\chi^2 = 3.56$, $p < .05$ (one-tailed). 18 out of 26 subjects showed the opposite preference to look at the false-belief window first in the LD condition, $\chi^2 = 3.85$, $p = .05$. The differences between first look in LD and TB conditions were significant, $\chi^2 = 5.76$, $p < .05$. They didn’t show any preference by first look in the HD condition, with 13 out of 26 subjects looked at the false belief window first (see Fig.4).
Fig. 4 Percentage of children’s first look to different windows. In both LD and TB conditions, the majority of children shifted their eye gaze to the belief-congruent window first. In the HD condition, however, they didn’t show any preference by first look.

*: $p < .05$ (one-tailed).

**Discussion**

In the TB task, preschoolers looked significantly longer at the side consistent with the actor’s true-belief, and they showed an opposite preference in the LD task in which they looked longer at the window in line with the actor’s false belief. However, in the HD task, they spent equal amount of time looking at both windows, indicating ambiguous anticipation of the actor’s behavior.
The differences between children’s anticipation in LD and HD tasks suggested that two- to-three-year-olds’ implicit eye gaze was sensitive to the processing demand of the tasks. The implicit belief reasoning measured by anticipatory eye gaze may not be demand free, and it requires efficient inhibition and selection process for success. This picture is compatible with a single system underlying belief reasoning, and the difference between verbal and non-verbal measurements is not qualitative but quantitative. The non-verbal tasks may require fewer inhibition and selection resources.

However, it’s also possible that the implicit system is distinctive from the explicit one, and they share a similar property – the sensitivity to processing demands of the tasks.

To separate these two hypotheses and to better understand the nature of the theory-of-mind reasoning mechanism, investigation of the mature belief reasoning in adults would help to address the question. It is not controversial that they possess the concept of belief and could pass the standard verbal tasks. Will more mature inhibitory processes help to facilitate their implicit eye gaze in such belief attribution tasks?

If there is a single system underlying both the verbal and non-verbal theory-of-mind reasoning, and the non-verbal task may free up more processing resources such that young infants could pass the low demand nonverbal task, adults should not have problem to pass the high-demand version of the non-verbal task with eye gaze. However, if there are two systems involved in theory-of-mind reasoning, as suggested by the dual-system account, the demand profile found in young children in experiment 1 may be also present in normal adults. What’s more we can ask adults verbal questions after the non-verbal measurements and compare these two measurements within-subjects. This verbal vs. non-
verbal comparison could further shed light on the potential difference/similarity between non-verbal/implicit and verbal/explicit reasoning.

**Experiment 2**

**Method**

89 subjects participated in the study (39 females), with a mean age of 19.5 years, ranged from 17.0-28.0 years. They were randomly assigned to the low demand (LD) false-belief (24), high demand (HD) false-belief (32), and true-belief (TB) (33) conditions. Subjects were undergraduate students from Rutgers University, and they earned course credits for participation. Another 65 subjects were excluded due to failure to reach attentiveness criteria (refer to Experiment 1). That is, we calculated subjects’ total looking time to both windows that may indicate their belief attribution. The total looking time was required to be above one second out of the last four seconds of anticipation to be included.

26 of the subjects attended the study during the piloting phrase, and they received both a false belief test (either LD or HD) and a true belief test, with the false belief test first. Therefore, we had 33 subjects received the true belief condition as between-subjects design, and another 26 subjects received the true belief condition as with-in subjects design. t-test revealed that there were no significant differences between their DLS in the true belief condition, $t (57) = 0.70, p = \text{n.s.}$ (between-subjects, DLS = -0.21; within-subjects, DLS = -0.32. All of them preferred looking at the true-belief window).

Therefore, we pooled over subjects’ eye gaze in the true belief test by treating all of them
as between-subjects results (more conservative than within-subjects results) in the following analysis, resulting in 59 individual data in the TB condition.

Unless stated, the methods were the same as used with children in Experiment 1.

Subjects were seated 60 to 80 cm away from the monitor (adjusted individually according to the tracking quality in calibration), in a separate booth enclosed with a curtain. Before the experiment, a 5-point circle calibration was applied. In the end of the study, after the four seconds pause, adults were asked the standard prediction question: “where will she look for the ball”. Their responses were recorded as Left box or Right box, and further coded as 0 (inconsistent with the actor’s belief) or 1 (consistent with the actor’s belief).

Results

1. Non-verbal performances: Differential Looking Score

Adults’ nonverbal performances were measured with reference to the Differential Looking Score (DLS) towards the false-belief window.

In terms of the screening criteria of subjects’ eye gaze in the last familiarization trial, we used it, either Fam-FL or Fam-PL, as a between-subjects variable to explore its influence on adults’ performance in the Belief test. Preliminary analysis suggested that the main effect of Fam-PL was not significant on subjects’ belief reasoning in the test trial (according to 2 by 3 ANOVA, the main effect of Fam-PL revealed a $p = .781$, no interaction), but the main effect of Fam-FL was significant (according to 2 by 3 ANOVA, the main effect of Fam-FL showed a $p < .05$, no interaction). Further t-test showed that the effect of Fam-FL was only present in the LD condition: those who made correct
anticipation by first look in the last familiarization trial (passed Fam-FL) showed stronger preference for looking longer at the false-belief window in LD than those who failed (t-test showed a $p < .01$), but the effect of Fam-FL was not significant in TB or HD conditions (TB, $p = .985$; HD, $p = .487$). Since adults surely know the contingency between the illumination and the actor’s subsequent behavior, and including subjects who failed the Fam-FL only made the results more conservative (especially in LD), we did not use this criterion with adults, resulting in a 3 (true belief, LD-false belief, HD-false belief) by 2 (side) ANOVA analysis on subjects’ nonverbal performances.

The 3 (task demand) by 2 (side) ANOVA showed that only the main effect of task demand was significant, $F(2, 109) = 10.15, p < .001, \eta^2 = .157$. Follow-up t-tests showed that subjects’ DLS were significantly different between TB and LD, TB and HD, and LD and HD (Fig.5). In the True-belief condition, they showed a significant preference for looking at the true-belief window (mean DLS = -0.26, $t(58) = -3.22, p = .002$), and that was completely opposite to their preference in the LD condition (mean DLS = 0.38 favoring the false-belief window, $t(23) = 3.54, p = .002$) (TB vs LD, $t(81) = 4.44, p < .001$). Performance was significantly different from TB in the HD condition, in which subjects showed no preference and looked equally across true- and false-belief windows (mean DLS = 0.11, $p = \text{n.s.}$) (TB vs HD: $t(89) = 2.72, p = .008$). The ambiguous preference in HD condition was also significantly different from LD condition, $t(54) = 1.69, p < .05$ (one-tailed) (see Fig.5).
Fig. 5 Mean of adult’s Differential Looking Scores. In both LD and TB condition, adults preferred for looking longer at the belief-congruent window. In the HD condition, they spent equal amount of time looking at belief-congruent and belief-incongruent windows.  

*: $p < .05$ (one-tailed); **: $p < .01$. 

In terms of subjects’ first saccades after the illumination, it failed to demonstrate adults’ anticipatory abilities in the present study: the percentage of subjects who correctly anticipated the actor’s subsequent behavior by first eye gaze was 63% in the LD condition (15 out of 24), 61% in the TB condition (36 out of 59), and 50% (16 out of 32). None of them was significantly different from chance. Therefore, we calculated adults’ eye gaze pattern instead, that is, the number of subjects whose DLS was above 0 in each condition. In LD condition, 19 out of 24 subjects’ DLS was above 0 (79%), suggesting that the majority of the subjects had a reliable preference for looking at the false-belief
window ($\chi^2 = 8.17, p = .004$). In the TB condition, the majority of the subjects (37 out of 55) preferred to look longer at the true-belief window (67%, $p = .010$). However, there was no significant preference in the HD condition (20 out of 32, 63%, $p = \text{n.s.}$). The observed differences between subjects’ eye gaze in TB and LD were significant ($\chi^2 = 12.67, p < .001$), and the differences were also significant between LD and HD ($\chi^2 = 3.81, p < .05$, one-tailed) (Fig.6).

![Diagram showing eye gaze preferences across conditions](image)

Fig.6 Percent of adults’ DLS favoring different sides. The majority of adults in LD and TB conditions preferred for looking longer at the belief-congruent window, but this pattern was not shown in HD condition.

*: $p < .05$, one-tailed; **: $p \leq .01$. 
2. Verbal vs. Non-verbal measurements

In general, 91% subjects answered the standard verbal question correctly. 23 out of 24 subjects (96%) in the LD and 28 out of 32 subjects (88%) in the HD said that the actor would search in the false-belief location, and 54 out of 59 subjects (92%) said she would search in the true-belief location in the TB task. All were significantly above chance ($p < .001$). McNemar test with paired samples revealed that their verbal responses were significantly different from their non-verbal looking performances (DLS pattern) in the HD false-belief task ($\chi^2 = 4.90, p < .05$) and the true-belief task ($\chi^2 = 6.86, p < .01$). That is, subjects were more likely to correctly predict the actor’s searching behavior in words than eyes (looking longer at the belief-congruent locations). But their verbal response did not exceed their nonverbal anticipations, and both revealed adults correct attribution of false belief to the actor (Fig. 7).
Fig. 7 Percent of adults’ nonverbal and verbal responses. Adults’ verbal responses to the standard questions were ceiling in all three conditions, better than their non-verbal responses with anticipatory eye gaze (DLS, above 0 when they preferred for looking longer at the false-belief side).

*: $p < .05$; **: $p < .01$.

**Discussion**

Adults spontaneously anticipated an actor’s behavior in accordance with her belief: in both the low-demand false belief task and the true-belief task, adults showed a clear and significant bias to look at the locations congruent with the actor’s belief. This result was consistent with previous findings (Senju, et al., 2009). However, they did not show the looking bias if the actor held a false belief that the toy was hidden in the currently empty
box while it was in fact hidden in the other one on the scene: in the high-demand false belief task, adults looked equally at the belief-congruent and belief-incongruent locations, making no clear predictions of the actor’s behavior. By contrast, adults’ verbal performances revealed a different picture: when asked to explicitly predict the actor’s behavior, adults’ performances were ceiling in all three conditions. The discrepancy between their verbal and non-verbal performances lends support to the dual systems hypothesis (Leslie, Friedman, & German, 2004).

**General Discussion**

We started from the question of whether the implicit belief reasoning can operate in the absence of an inhibition and selection process. Specifically, we were interested in whether subjects’ anticipatory eye gaze is sensitive to the processing demand of the task. The effects of processing demands on younger children’s explicit belief understandings are well documented (e.g. Leslie et al., 2005), however, it was unknown whether the implicit reasoning system is also subjected to task demands. Built on the findings that young children’s explicit performances could be facilitated when the desired target was removed from the scene (Bartsch, 1996; Wellman, et al., 2001), we manipulated the demands of the non-verbal anticipatory task (Southgate, et al., 2007) by keeping the target on scene (HD) or completely removing it (LD). Subjects’ nonverbal performances turned out to be better in the LD condition: they showed strong bias to look at the locations congruent with the actor’s false belief. Nevertheless, adults did not have any preferences in the HD condition, looking randomly at the locations in line with the
actor’s false belief and locations consistent with a true belief. Even worse, preschoolers’ eye gaze in the HD condition was not different from their performance in the TB condition, completely incongruent with the actor’s false belief. Together, these results suggested that spontaneous belief attributions measured by the nonverbal anticipatory task were sensitive to the processing demands of the tasks.

1. ToMM + selection process and the properties of the dual systems

Recall that we proposed two possible pictures at the beginning. It could be that the implicit system does not depend on processing demand: subjects either understand the story of the task and reason about belief correctly, or fail to interpret the story at all. The variations of the processing demands, which were reported to affect children’s explicit performances, won’t have similar effects on their implicit reasoning. This hypothesis was dismissed given the observed demand profile in adults’ and two-to-three-year-olds’ implicit anticipations.

By contrast, subjects’ better performances in the low-demand condition and ambiguous or incorrect performances in the high-demand condition were compatible with the other hypothesis: the implicit system was more like the explicit one than was previously thought. Both are subject to the processing demands of the tasks. Therefore, the theory-of-mind mechanism would give rise to the representation of appropriate belief contents, but that’s not sufficient for correct belief reasoning; the selection process is needed to apply our processing resources to make the correct belief contents potent for inferences (Roth & Leslie, 1998). From this perspective, the implicit and explicit reasoning systems
are uniform, and they have similar properties in terms of their sensitivities to task
demands.

In light of the similarities with the two systems, could we go further to claim that there
are no dual-systems underlying theory of mind abilities, and the distinctions are caused
by the verbal or non-verbal nature of the tasks? The discrepancies between adults’ verbal
and nonverbal performances in the present study cast doubt on the single-system
hypothesis. Although adults correctly anticipated the actor’s behavior with implicit eye
gaze in the LD and TB condition, they did that even better in words explicitly. What’s
more, their eyes failed to show their false belief understanding in the HD condition,
strikingly contrary to their ceiling performances in the voluntary verbal task. Combined
with younger children’s performance in verbal and nonverbal tasks, these results reveal a
double-dissociated picture: younger children who typically failed the verbal belief tasks
could demonstrate their belief reasoning competence implicitly (Clements & Perner,
1994; Garnham & Perner, 2001; Garnham & Ruffman, 2001); and adults in the present
study were generally better at reasoning about false belief explicitly rather than
implicitly.

These discrepancies are consistent with the dual systems model. On the one hand, the
theory-of-mind mechanism, by its modular nature, could generate the correct belief
contents, but its potency may only be sufficient for children to understand belief correctly
in some limited contexts (e.g., the nonverbal context and the low demand context). With
limited processing resources, younger children won’t be able to show their belief
reasoning competence explicitly in words, and thus tend to do better in the nonverbal
tasks. On the other hand, after four years of age, children have enough resources to
reason about belief explicitly in the standard tasks. With further development, adults could reason about false belief in more complex situations that impose even higher processing demands (Cassidy, 1998; Friedman & Leslie, 2004a, 2004b; Leslie, German, & Polizzi, 2005; Leslie & Polizzi, 1998). However, the competence directly provided by the module remains relatively unchanged and limited throughout development, while people’s explicit capacities come to surpass their implicit competence. That’s why adults were better at explicit reasoning in the present study.

Despite similarities, the implicit and explicit systems are distinct, operating separately and following different development trajectories. Whereas the implicit system may be able to apply some rudimentary processing resources, it can’t take advantage of the later developing processing capacities; the explicit system can make use of the increasing power of the selection process, and thus manages to apply the belief representation to a broader range of situations.

2. VOE vs. AL discrepancy

In previous studies, infants’ nonverbal belief competence has been shown by both anticipatory eye gaze measures and by looking time in violation-of-expectation (VOE) studies (Onishi & Baillargeon, 2005; Scott & Baillargeon, 2009; Scott et al., 2010; Song & Baillargeon, 2008; Surian, Caldi, & Sperber, 2007; for a review, see Baillargeon, Scott, & He, 2010). Notice that in the VOE version of the Sally-Anne task, infants were shown the high-demand condition in which the target remained on the scene. Contrasting with poor performance in the present HD condition, even 15-month-olds looked reliably longer when the actor with a false belief reached for the toy’s real location, suggesting
that they correctly expected the actor to search for the other location consistent with her false belief (Onishi & Baillargeon, 2005). How do we explain the demand profile of the implicit system observed here, and the absence of sensitivity to task demands in the VOE studies?

One possibility is that the discrepancy is caused by the differences between VOE and Anticipatory Looking tasks. Although neither of them requires any structural responses from subjects, in the VOE task, subjects simply watch the story unfold, and they are shown the outcomes directly instead of calculating them; in the AL task, subjects have to predict these outcomes given the false belief context. In both cases, ToMM would compute the representations of the agent’s belief automatically, by integrating different pieces of information from the context. However, at least one extra step of computing is required to make anticipations, namely, the selection of appropriate action representation. Therefore, VOE tasks may require even less computing resources beyond the belief representations. Consequently, infants may be able to marshal the extra demands imposed by high-demand tasks, showing no sensitivity to the demands generated by the hidden object on the scene. Similar evidence was found in verbal tasks, where showing the outcomes could enhance three-year-olds explicit performances (Bartsch & Wellman, 1989; Robinson & Mitchell, 1995). For example, Bartsch and Wellman (1989) found that three-year-olds were able to use the character’s false belief to offer explanations to the character’s behavior, but they constantly failed to predict his behavior out of false belief.

Therefore, we may be able to find other situations requiring even higher processing demands that may evoke the demand profile measured by VOE tasks. Implicated by verbal studies, a false belief task involving a desire to avoid a particular location may be
a good place to start (Cassidy, 1998; Friedman & Leslie, 2004a, 2004b; Leslie, German, & Polizzi, 2005; Leslie & Polizzi, 1998). For example, Leslie & Polizzi (1998) told four-and-five-year-olds that Renee wanted to put a piece of fish into one of two boxes. But there was a sick kitten in one box, and Renee did NOT want to make the kitten worse by giving her more food, therefore, she did NOT want to put the fish in the box with the kitten. But when she was away to get the fish, the kitten crawled into the other box. Then, Renee came back with the fish, and children were asked both a “think” question (“where does Renee think the kitten is”) and a “prediction” question (“which box will she go to with the fish?”). Leslie and colleagues found that only 38% of four-year-olds appropriately predicted the girl’s behavior from a false belief with an avoidance desire, even though all subjects correctly answered the think question (Leslie & Polizzi, 1998). Thus, poor performance in the prediction question was not caused by a failure to understand false belief, but a failure to marshal double inhibitions imposed by false belief and avoidance desire. We are currently working on a nonverbal version of the false belief tasks associated with avoidance desires. What’s more, by comparing subjects’ verbal and non-verbal performance in these tasks, we may be able to reveal when the explicit system first outstrip the implicit one.

3. Limitations and remaining questions

Children’ and adults’ poor performances in the HD condition were different from the anticipatory results measured by the “I wonder” prompt (Clements & Perner, 1994; Ruffman, Garnham, Import, & Connolly, 2001). In Clements & Perner’s (1994) study, they used a HD version false belief task with the “I wonder” anticipatory looking measurement, and found that two-to-two and a half-year-olds failed to anticipate the
actor’s behavior based on her false belief, just like the subjects’ performances in the present study. Children older than 2 years and 11 months passed the prompted HD false belief task with their anticipatory eye gaze. It is not clear what factors enhanced the preschoolers’ performances in the verbal prompt study, or what factors may hinder children’s anticipatory eye gaze in the present nonverbal study with Tobii eye tracker.

**Conclusion**

The present study investigated adults’ and preschoolers’ belief attribution with the anticipatory looking method, and we found that preschoolers’ and adults’ eye gaze correctly anticipated an actor’s behavior in line with her belief in both LD and TB tasks. However, adults’ predictions were better with verbal responses, suggesting that the implicit belief reasoning system is independent from the voluntary system and it stays relatively constant across ages. More importantly, subjects’ nonverbal performances with their eyes were better in LD condition than in the HD condition, implying that the implicit belief system is sensitive to the processing demands of tasks, similar to the explicit system. Combined with prior findings in verbal tasks, we concluded that there are dual systems involved in belief reasoning, and the two systems share the same property in terms of their sensitivity to task demands. It remains to be seen how the dual systems theory involving the ToMM/SP model can account for these and still on-going studies.
References


Footnotes

1. The implicit system is usually measured by non-verbal tasks and the explicit reasoning is measured in verbal tasks, therefore, implicit/explicit and non-verbal/verbal were used interchangeable here.