The interpretations of infants' non-verbal responses in violation-of-expectation (VOE) false belief scenarios are subject to intense theoretical debate. In Experiment 1, adults provided online narratives for VOE scenarios meant to tap understanding of false beliefs about object location, perception and identity. Adults provided cognitively-oriented narratives for the location scenario when instructed to track beliefs and, for this scenario only, participants evaluated the unexpected outcome as unexpected and the expected outcome as expected. Adults had mixed views about the perception scenario, and judged the identity scenario where the agent acted in violation of his belief as being reasonable. Experiment 2 confirmed that when the identity scenario was turned into an action task that was time-pressured, adults failed to act in a belief-based manner. We should be cautious in drawing firm conclusions about mentalizing in infancy when adults' narratives and estimates of the expectedness of outcome events suggest that only certain VOE scenarios were interpreted in their intended fashion.

KEYWORDS: False Belief; Violation-of-Expectation; Adults; Interpretations; Action

## 1. Introduction

Decades of research on standard false belief tasks requiring verbal reasoning, indicate that theory-of-mind (TOM; also referred to as mindreading) emerges in humans from about 4 years of age (Wellman, Cross, & Watson, 2001). This basic developmental trend is illustrated in Wimmer and Perner's (1983) object location false belief task, in which Maxi is shown storing his chocolate in a cupboard and leaving the scene. During Maxi's absence, his mother transfers the chocolate to a table drawer. Children are directly asked to predict where Maxi will go to look for his chocolate. Most 3-year-olds incorrectly answer that Maxi will look in the actual location rather than in the place he left it. From age 4, children typically arrive at the correct answer, appreciating that others may have false beliefs about an object's location. From this age TOM is conceptually unified: the emerging understanding of belief includes not just the possibility for beliefs to misrepresent the world, but that people's beliefs can be about specific aspects of things (Rakoczy, Bergfeld, Schwarz, & Fizke, 2015). The classical view is that there is deep conceptual change in children's belief reasoning over the preschool years, and advances in language, executive function and participation in complex and diverse social interactions help children learn about subjective mental perspectives (Apperly, 2011; Low & Perner, 2012; Perner, 1991).

A challenge to the conceptual shift account comes from research using the violation-of-expectation (VOE) technique where the amount of time infants look at different stimuli is measured. Onishi and Baillargeon (2005) showed 15-month-olds scenarios of an agent forming either a true belief or false belief about an object's location. The agent was ultimately shown searching in the belief-compatible or the belief-incompatible location for the target object. Infants looked longer when the agent searched in a location that did not match her belief. Onishi and Baillargeon interpreted the longer looking as evidence of infants understanding that others act on the basis of their beliefs and that these beliefs are representations that may or may not correspond with reality.

In a follow-up study, Song and Baillargeon (2008) showed 14.5-month-olds an event sequence where an agent displayed a preference to reach for a black skunk instead of a blue-haired doll (skunk condition; toys' locations were counterbalanced). In the agent's absence, the skunk was put into a plain box and the doll was put into a box which had a black tail attached to the inner edge of its lid.

The agent returned and was shown either reaching for the tail box or the plain box. Infants looked longer when the agent was shown reaching for the plain box, suggesting that 14.5-month-olds can track others' false beliefs based on the way an object appears. Other studies suggest that infants can even causally and systematically reason about others' beliefs. Scott and Baillargeon (2009) showed 18-month-olds an agent who was presented with a 2-piece penguin toy and a solid penguin toy. The agent repeatedly took out a key and hid it inside the 2-piece toy. When the agent was away, the 2-piece toy was assembled (to match the appearance of the one-piece penguin) and a transparent box was placed over it. An opaque box was placed over the indivisible toy so that the agent, on her return, would mistakenly believe that the observable penguin was the indivisible one. Then the agent, with key in hand, was shown to either reach for the transparent box or the opaque box. Infants looked longer when the agent reached for the transparent box than when the agent reached for the opaque box. The pattern of looking responses was taken to show that infants: inferred the agent's goal of hiding the key in the divisible toy; attributed that the agent falsely believed that the toy in the transparent box was the indivisible toy (when it was really the divisible toy); falsely believed that the divisible toy was located in the opaque box (when it was really in the transparent box); and expected the agent to reach for the opaque box and were surprised when she reached for the transparent box.

How can infants display sensitivity to others' false beliefs when responding in some ways while 3-year-olds treat false belief as impossible when responding to the very same situation in other ways? Baillargeon and colleagues (2010) explain that VOE techniques only tap the early developing psychological reasoning system and infants can simply express their abstract understanding of an agent's mental states as they observe a scene unfold. By contrast, standard TOM tasks make demands on language and executive function skills that develop slowly toward the end of the preschool years. These additional demands mask 3-year-olds' conceptual competency. There are, however, significant challenges to Baillargeon's argument (Sodian, 2011, 2016). First, children's executive function is correlated uniformly with performance on TOM tasks that impose high or low executive demands (Carlson, Claxton, & Moses, 2015). Moreover, preschoolers with a high level of executive function skills pass standard false belief tasks at the same time as their control counterparts (Sabbagh et al., 2006). Second, infants' responses to VOE tasks may just as well be explained by differential attention

to the perceptual novelty of certain test outcomes with respect to previously encoded events (Heyes, 2014) or to the following of stimulus-response behavior rules (Ruffman, 2014). Finally, the combined evidence of success amongst infants in diverse VOE studies is largely based on data between age groups and between children, making it difficult to compare depth of understanding on indirect versus direct measures of false belief reasoning (Yott & Poulin-Dubois, 2016).

VOE methodology is primarily designed to answer questions about infants' basic sensory discriminations (Haith, 1998); higher-level explanations of how the unexpected and expected event sequences are meant to be interpreted in the VOE studies are provided by the researchers themselves. Scott and Baillargeon (2009) construe their VOE scenario as engaging infants' understanding of false beliefs about object identity. Others disagree with the interpretation (e.g., Butterfill & Apperly, 2013; Fizke, Butterfill, & Rakoczy, 2013; Low, Apperly, Butterfill, & Rakoczy, 2016; Ruffman, 2014; Wellman, 2014; Zawidzki, 2013) and suggested that the scenario only taps tracking of others' false beliefs involving the types of objects present and not object identity. An agent's false belief over object identity, in the strict numerical sense, would result in mistakes involving expansion (agent thinks that there are two distinct objects when there is only one) or mistakes involving compression (agent thinks that there is only one object when there are two). Butterfill and Apperly argued that in the case of Scott and Baillargeon's task, there were always two objects present on the stage (the divisible penguin and the indivisible penguin) and the agent possessed the same knowledge. Infants could have passed the task by tracking the agent's false belief of the types of objects present: the object on that side is a divisible toy so the object on the other side is an indivisible toy.

To shed light on the debate over the interpretation of infants' performance in VOE studies, the current research asks whether adult participants would explain VOE event scenarios in the same way that was intended by the task developers themselves. If adults find it difficult to grasp the event sequences that make up complex VOE tasks, it would be difficult to sustain a rich mentalistic interpretation of infants' looking behavior on the same tasks. In Experiment 1, we examined adults' explanatory narrations for false belief VOE scenarios that have been used in three prominent infant studies. In Experiment 2, we measured adults' non-verbal reaching actions to assess the construct validity a VOE scenario that has been suggested to involve false beliefs about object identity.

## 2. Experiment 1

### *2.1. Method*

In Experiment 1 adults provided online narratives for three VOE films that have been suggested to involve an agent's false belief about: object location (Onishi & Baillargeon, 2005), object perception (Song & Baillargeon, 2008) and object identity (Scott & Baillargeon, 2009). At the end of each film, participants saw either an expected ending (the agent acted in a way that was consistent with his false belief) or an unexpected ending (the agent acted in a way that was inconsistent with his false belief). We expected that adults would be able to explain and make sense of the object location and object perception false belief VOE scenarios, but that there would be ambiguity over the meaning of the object identity VOE scenario.

We also compared adults' narrations and judgments depending on whether participants were instructed to watch the VOE films freely, to track the objects in the films, or to track the beliefs or perspectives of the displayed agent. According to Apperly and Butterfill (2009), belief inferences make deep and lasting demands on executive resources, even in adults, and are not likely to be automatic. Moreover, belief inferences can be computationally intractable without constraints on which aspects of others' situation might be relevant for interpreting or predicting their behaviors (Apperly, 2011). If it turns out that adults, from the start, are automatically set to make inferences about an agent's belief, perspective or motivational state whenever they see an agent behave in some way, then adults' narratives should make references to agents' internal cognitions even when participants are not specifically instructed to track mental states. Following Apperly and Butterfill, however, we predicted that adults would only spontaneously make references to internal cognitions in their narratives when they were explicitly instructed to track others' internal mental states.

#### *2.1.1. Participants*

A total of 72 adults (undergraduates: 17 males and 55 females) ($M$ = 18.92 years; $SD$ = 2.83; Range = 17 to 40) participated in exchange for partial course credits. The sample size aligns with other studies that have examined adults' narratives or verbal explanations of the reasons behind others' actions (e.g., Abel, Happé, & Frith, 2000; Lagattuta, 2005; Lovett & Pillow, 2010).

#### *2.1.2. Materials*

The object location film (Figure 1a) was adapted from the false belief green condition of Onishi and Baillargeon's (2005) study. It started with a familiarization phase that comprised three trials. In trial 1, a watermelon toy was positioned on a table between a green box (agent's left) and a yellow box (agent's right). When the curtain was raised, the agent picked up the toy and placed it into the green box. In trials 2 and 3, when the curtain was raised, the agent placed his hand into the green box where the toy was located. In the belief-induction event, while the agent was away, the toy silently moved from the green box to the yellow box. For the expected ending, the agent acted in accordance with his false belief and reached into the green box where he thought the toy was located. For the unexpected outcome, the agent reached into the yellow box where the toy was really located.

*** Figure 1 ***

The object perception film (Figure 1b) was adapted from Song and Baillargeon's (2008) false belief condition. There were four familiarization trials. For trials 1 and 2, the agent sat before a blue-haired doll (on the agent's right side) and a black skunk (on the agent's left side). A disembodied hand entered the stage and placed the skunk on top of a mat, and the doll on top of another mat. The agent grabbed the skunk after the hand exited the stage. The sequence was the same for familiarization trials 3 and 4, except that the skunk was on the agent's right side and the doll was on the agent's left side, and that the hand placed the toys on shallow containers. The belief-induction event was as follows. Whilst the agent was away, the stage floor presented two boxes: a plain box and a tail box (so-called because a skunk's tail was attached to the inner edge of the box's lid). A disembodied hand entered the stage and lifted the lid of the tail box five times and then lifted the lid of the plain box five times. Then the skunk appeared in front of the plain box and the doll appeared in front of the tail box. The disembodied hand entered the stage and deposited the doll into the tail box and the skunk into the plain box. For the expected outcome, when the curtain was raised, the agent acted in accordance with his false belief of the way things appeared and reached for the tail box. For the unexpected outcome, the agent reached for the plain box.

The object identity film (Figure 1c) was adapted from Scott and Baillargeon's (2009) false belief condition. In familiarization trials 1 and 2, the agent faced two penguin toys, a divisible penguin (the head piece sat beside the body piece) and an indivisible penguin (solid toy). A disembodied hand

entered the stage, lifted the indivisible penguin by its head and placed it on top of a stand. Then the hand picked up the body piece of the divisible penguin and placed it on top of the other stand. The hand then picked up the head piece of the divisible penguin and placed it beside its body part on the stand. The agent then raised his right hand to reveal a key (he waved the key several times), placed the key into the body piece of the divisible penguin, and placed the head piece on top of the body piece (once assembled, the divisible type of toy looked like the indivisible type of toy). Familiarization trials 3 and 4 were the same except that the divisible penguin was on the agent's left side and the indivisible penguin was on the agent's right side, and that there were two trays instead of two stands. The belief-induction event was as follows. In the agent's absence, the stage presented two boxes, one was transparent and the other was opaque. The divisible penguin sat in front of the transparent box whilst the indivisible penguin sat in front of the opaque box. A disembodied hand entered the stage and placed the indivisible penguin flush against the opaque box. The hand then placed the body of the divisible penguin flush against the transparent box and assembled the head piece on top of the body piece. The hand then covered the assembled divisible toy with the transparent box, and covered the indivisible toy with the opaque box. For the expected outcome, when the curtain was raised, the agent, reached for the opaque box, believing (according to Scott and Baillargeon's rationale) that there was an indivisible toy underneath the transparent box. For the unexpected outcome, the agent reached for the transparent cover. The event components making up each film are in Supplementary Materials.

*2.1.3. Procedure*

Participants were individually tested in a quiet room. They all watched three types of false belief VOE films (order counterbalanced); half of them were exposed to scenarios with unexpected outcomes, and half to scenarios with expected outcomes. Participants were informed that that they would see three films whereupon they had to provide story narrations for each one. They were told that each film would be shown twice, with the second screening immediately following the first screening. They were told to use the first screening as a previewing opportunity to silently consider what they might say when story-narrating the film during the second viewing. Previewing event-based sequences assists story processing and encourages more story-like narratives (Shapiro & Hudson, 1991) and does not diminish story engagement or enjoyment (Leavitt & Christenfeld, 2011).

Participants were assigned to receive one of three types of instructions: neutral, object tracking or mental state tracking. Participants in the neutral group were given the following instruction: "Your story narration should explain *the events* in the film to the best of your abilities." Participants in the object tracking group were given the instruction: "Your story narration should explain *the location of the toy or toys* in the film to the best of your abilities." Participants in the mental state tracking group were given the instruction: "Your story narration should explain *the motives and beliefs or perspectives of the actors* in the film to the best of your abilities." The films were played on a 19-inch television and participants spoke into a microphone. In an exit questionnaire, participants responded to an open-ended question ("What did you make of the ending?") regarding each film's conclusion (they were prompted by still frames depicting the ending of each VOE scenario).

*2.2. Results and Discussion*

Analyses focused on the context of the narrative generated for the belief-induction event (scores ranged from 0 to 2) and the ending event (scores ranged from 0 to 2) of the different films. A score of 0 was given if the participant offered no narration for the belief-induction or ending events, or was substantially off-topic. A score of 1 was given if the narrative for that event was purely descriptive in context. A score of 2 was given if the narrative included any cognitive information that made explicit reference to an internal mental state (see Supplementary Materials for examples). Two raters independently coded 25% of the narratives with 96% reliability, and one of the raters coded the remaining narratives alone. Summing across the belief-induction and ending events, each participant could, for each film scenario, receive a narrative context score that ranged from 0 to 4 points. The higher the context score, the more cognitively oriented the narrative produced.

*** Table 1 ***

Participants' narrative context scores (means reported in Table 1) were submitted to a mixed model 3 (scenario: location, perception and identity) x 3 (instruction: neutral, object, or belief) x 2 (outcome: unexpected or expected) ANOVA with scenario as a within-subjects variable and instruction and outcome as between-subjects variables. The 3-way scenario x instruction x outcome interaction was significant ($F(4, 132) = 3.06$, $p = 0.019$, $\eta_p^2 = 0.085$). The 3-way interaction reflected the presence of 2-way interactions between instruction and outcome for the location scenario ($F(2, 66)$

= 4.00, $p = 0.023$, $\eta_p^2 = 0.108$) and the perception scenario ($F(2, 66) = 4.68$, $p = 0.013$, $\eta_p^2 = 0.124$). In the location scenario, and only with a belief-focused instruction, participants' narrative of the unexpected outcome was more cognitively oriented than participants' narrative of the expected outcome ($F(1, 22) = 5.97$, $p = 0.023$, $\eta_p^2 = 0.213$). For the perception scenario, when given an object-focused instruction, participants' narrative context score for the expected outcome was higher than participants' narrative context score for the unexpected outcome ($F(1, 22) = 6.82$, $p = 0.016$, $\eta_p^2 = 0.237$). The 2-way instruction by outcome interaction for the identity scenario was not significant ($F(2, 66) = 0.04$, $p = 0.962$); there was only a significant main effect of instruction ($F(2, 66) = 8.57$, $p < 0.001$, $\eta_p^2 = 0.206$) whereby participants who received the belief-focused instruction demonstrated a higher narrative context score than participants who received the object-focused instruction.

The narrative context scoring does not provide insight into what participants themselves think about the ending events of the VOE scenarios. Consider a participant who narrated the ending of the location scenario as, "And then it [the screen] opens, and the man reaches into the yellow box like he knew it had been moved." Whilst the narration, in mentioning an internal mental state (i.e., "knew"), would receive a context score of 2 points, the statement does not illuminate whether or not participants themselves deemed the actor's final action as being expected or unexpected, justified or not justified. We were able to gain better insight into participants' judgments about the intelligibility of the various endings by analyzing their responses on the exit questionnaire. Participants' judgments about the endings were coded as a category variable comprising of three levels: 0 = ending was deemed expected; 1 = ending was deemed unexpected; or 2 = don't know or ambiguous response (examples are provided in Supplementary Materials). Two raters independently classified 25% of participants' judgments with 92% reliability and one of the raters coded the remaining judgments alone. Loglinear analyses were conducted to examine the relationship between the categorical variables of instruction, outcome and own judgment.

For the location scenario, parameter estimates from a loglinear analysis indicated that there was only a significant 2-way outcome by judgment interaction ($Z = -5.27$, $p < 0.001$, 95% CI = -2.32 to -1.06). Regardless of instruction, a majority of participants who viewed the unexpected outcome themselves judged the ending as being unexpected ($n = 29/36 = 81\%$) whilst all of the participants

who viewed the expected outcome themselves judged the ending as being expected ($X^2(2) = 68.11$, $p < 0.001$). For the perception scenario, there was only a significant 2-way outcome x judgment interaction ($Z = -2.38$, $p = 0.017$, 95% CI = -0.89 to -0.09). Difference in $Z$-scores suggested that the outcome by judgment interaction effect was weaker for the perception scenario ($|Z| = 2.38$) than the location scenario ($|Z| = 5.27$). The lower $Z$-score of the outcome by judgment interaction for the perception scenario indicated that although a majority of participants who watched the expected outcome also judged the ending as being expected (31/36 = 86%), participants who viewed the unexpected outcome were divided as to whether they themselves judged the ending as being unexpected ($n = 18/36 = 50\%$) or as being expected ($n = 16/36 = 44\%$) ($X^2(2) = 15.50$, $p < 0.001$). Of those 16 participants who judged the unexpected outcome of the perception scenario as being expected, 81% ($n = 13$) of them explained that the agent reaching for the plain box was a sensible thing to do because the tail box was too obvious a trick to fall for.

For the identity scenario, parameter estimates indicated that there was only a main effect of own judgment ($Z = 5.03$, $p < 0.001$, 95% CI = 0.53 to 1.21) whereby more than half of the judgments about the ending were of the expected variety ($n = 47/72 = 65\%$) ($X^2(2) = 34.75$, $p < 0.001$). Indeed, many participants judged the ending of the identity scenario as being expected regardless of whether they were watching the expected outcome ($n = 20/36 = 77\%$) or the unexpected outcome ($n = 27/36 = 75\%$). We further examined the 'expected' judgments offered by the 20 participants watching the expected outcome of the identity scenario: all of them made judgments based on the types of objects present rather than numerical identity (e.g., "Makes sense from what he experienced; there was always one doll without a lid on for him to put the key in, and he can see one of the dolls and it's got its lid on"). Amongst the 27 participants who judged the unexpected outcome of the identity scenario as being 'expected', 25 of them (93%) reasoned that the ending was sensible because the agent was reaching towards a box where one could actually see something (e.g., "He chose that box because he can see what's in it. I would probably choose the one that I could see what's in there. It kind of makes sense that he'd go for that one, because you can see what's in there").

Supporting our first hypothesis, adults perceived the event sequences of Onishi and Baillargeon's (2005) VOE task to be meaningful. Supporting our second hypothesis, participants only

spontaneously talked about the agent's cognitive state in this scenario when they were instructed to generate narratives that explained the motives and beliefs or perspectives of actors in the film. Where direct judgments were concerned, participants regardless of instruction interpreted the unexpected outcome of the object location scenario as being unexpected and judged the expected outcome of the object location scenario was being expected.

The results for the other VOE scenarios were less clear cut. Participants watching the expected outcome of the perception scenario generated narratives with higher context scores than compared to participants watching the unexpected outcome of the same scenario; this effect was observed when participants were instructed to focus on object whereabouts. An object-focused instruction may have motivated participants to pay more attention to the way that the tail box appears and the subjective reason for the agent selecting the tail box (with the undesired doll) rather than the plain box (with the desired skunk). However, whilst adults judged the expected outcome of the perception scenario as being expected, there was ambiguity over what was assumed to be an unexpected outcome. Adults were divided in their judgments about whether the unexpected outcome was expected or unexpected; many participants stated that it was logical for the agent to reach for the plain box as the tail box would be much too obvious a trick for someone to fall prey to.

The majority of narratives generated for the identity scenario were merely descriptive and participants did not make explicit reference to the agent's cognitive states, regardless of instruction manipulation. Surprisingly, the majority of adults viewing the unexpected outcome judged it as being expected; many stated that it was a reasonable thing to do, to reach for an object that one encounters in the transparent box rather than take a chance on something that may or may not be found underneath the opaque box. The most important finding, however, was that a majority of participants provided a converging explanation for why the agent reached for the opaque box: a divisible toy must be under the opaque box if the toy in the transparent box is an indivisible one. In this way, adult participants' judgments align with Butterfill and Apperly's (2013) minimalist interpretation of what Scott and Baillargeon's (2009) VOE task might be tapping into, that is, reasoning about *types* of object present rather than identity per se. In Experiment 2, we tested whether adults would conceivably reach for the transparent box if they happened to find themselves in the actor's shoes.

## 3. Experiment 2

### *3.1. Method*

The identity VOE scenario that participants passively watched in Experiment 1 was turned into an active participation task in Experiment 2. Across three familiarization trials, participants sat in front of an actor and observed the actor hide a key in the body of a divisible penguin toy. In the fourth familiarization trial, the actor left the scene and participants took on the role of the actor. Participants sat in the actor's chair and also had to place the key inside the body of the divisible toy and then reassemble the toy. The innovative component to this experiment was measuring how participants would react in the ending phase when the blind was raised and participants were faced with a transparent box (containing an assembled divisible toy) and an opaque box (containing an indivisible toy). In Experiment 1, participants suggested that it would be reasonable for an agent to reach for the transparent box with the visible penguin, and that they would have acted similarly if they were in the same situation. According to Apperly and Butterfill (2009), our ability to track object encounterings imposes relatively fewer demands on general information processing resources compared to making attributions about object perceptions (and beliefs) as standardly conceived; minimal mindreading inferences are more likely to be deployed in time pressured situations where cognitively efficient responses are critical. For Experiment 2, then, we contrasted participants' actions in a condition where there was urgency to the box selection as compared to a condition where there was no urgency. We hypothesized that there would be more first reaches to the transparent box when the task scenario was set in an urgent context than when the task was set in a non-urgent context. Analysis of response latencies can potentially illuminate the cognitive processes that underpin immediate actions (Atance, Bernstein, & Meltzoff, 2010; Edwards & Low, 2016). Following the view that attributing beliefs about objects, as compared to tracking encounterings of objects, is not automatic but imposes deep and lasting cognitive demands, we hypothesized that first reaches to the opaque box should be slower than first reaches to the transparent box.

#### *3.1.1. Participants*

A total of 32 adults (undergraduates) participated in exchange for partial course credits (8 males and 24 females; $M$ = 19.22 years, $SD$ = 2.04; range = 18 to 27). The sample size is in keeping with

other studies where adults' reaching actions have been measured as a way of evaluating behavioral data from infants (e.g., Hespos, Gredebäck, von Hofsten, & Spelke, 2009).

*3.1.2. Procedure*

Participants were individually tested in a quiet laboratory. A hidden video camera was mounted on the ceiling and recorded hand movements on the stage. Each participant was randomly assigned to either the urgent or the non-urgent condition. In both conditions the experimenter provided the following instructions, "When this blind is raised, you will see an actor behind it. Please attend to everything that is happening in front of you, until I give you further instructions. It is important that you pay attention to what is happening, as later you will be asked to do something based on your observations." The participant sat in front of the stage whilst the experimenter sat behind the participant. The blind was raised revealing an actor seated on the opposite side (facing participant) and familiarization trial 1 commenced (see Figure 2).

*** Figure 2 ***

In the urgent condition, a disembodied hand reached out from behind the side curtains with a key, placing it on the stage floor. The hand then brought out an indivisible penguin toy, and placed it on the left-side of the stage (the actor tracked all of the hand's movements). The hand re-appeared from behind the curtain and placed the head, then the body of a divisible penguin toy on the right-side of the stage. Once the hand retracted, an alarm was set off. This was a loud civil defence siren with countdown effect, played through speakers hidden underneath the stage (see https://www.youtube.com/watch?v=NMi_IJOsS1c). The aim was to contextualize the actor's action as being time-pressured. A civil defence siren was appropriate to our testing environment as all of our participants live on an island nation that is a hot bed of seismic activity, and emergency sirens are treated as demanding rapid action. The actor quickly grabbed the key, placed it into the body of the 2-piece penguin and placed the head back on. When the head of 2-piece penguin was placed back on, the alarm stopped. The blind then rolled back down marking the end of the trial. The sequence in familiarization trial 2 was similar to familiarization trial 1 except that the hand moved the divisible penguin first. The sequence in familiarization trial 3 was similar to familiarization trial 1 except that the divisible penguin was placed on the left-side of the stage. Before familiarization trial 4, the

experimenter gave the participant the following instructions, "Ok, now things are going to change. You're going to take a turn playing the actor behind the blind and I'm going to take your role as the observer." The experimenter brought the participant to the seat behind the stage (the actor by that time had left the scene) and said, "Remember you are now the actor and I'm the observer. Please act according to your previous observations." The experimenter then sat down where the participant had been originally seated. At that point, the blind was raised for familiarization trial 4, whereupon the following were placed on the stage: the key (centre-stage), the indivisible penguin (right-side) and the pieces of the divisible penguin (left-side). The siren sounded after the hand exited from the stage. All participants responded by grabbing the key, placing it into the body of the divisible penguin, and reassembling its head; at that point the siren alarm stopped and the blind was lowered. Across familiarization trials 1 to 3, the alarm could go off 3 seconds, 5 seconds or 9 seconds after the hand left the stage (timings across the four versions of the familiarization phase were: 3s-5s-9s or 5s-3s-9s or 9s-3s-5s or 9s-5s-3s). For familiarization trial 4, the alarm went off 7 seconds after the hand exited from the stage.

The test trial directly followed the familiarization phase. While the blind was lowered, occluding the participant's view (participant still acting as agent), the experimenter silently assembled the pieces of the divisible penguin on the left-side of the stage, and placed a transparent box over it. The experimenter placed the indivisible penguin toy on the right-side of the stage, and placed an opaque box over it (sides counterbalanced). The key was placed between the two boxes, then the screen was raised and the siren was immediately triggered. The variable of interest was the box that the participant-agent first touched or reached for. As in the familiarization trials, the siren was only deactivated when the participant inserted the key into the divisible penguin and reassembled it. At the end of the experiment, participants were asked to explain why they reached first for a particular box.

In the non-urgent condition, instead of a civil defence alarm being triggered, a non-emergency sound was used: a hotel call bell was tapped from behind the curtains, producing a single ding sound. The non-urgent condition was exactly the same as the urgent condition in all other respects.

*3.2. Results and Discussion*

There was a significant association between first reach and condition, $X^2(1) = 12.52$, $p < .001$. In the non-urgent condition, all 16 participants first reached for the opaque box, whereas in the urgent condition, only 7 did so (44%). In terms of explanations, almost two-thirds of the participants in the non-urgent condition (10/16) reasoned about object types (e.g., "that one [in transparent box] looked like it was solid and didn't have like a lid and I just assumed that [in opaque box] would be set up like previously"). The remaining 6 participants in the non-urgent condition could not explain their first reach decision. Of the 9 participants in the urgent condition who first reached for the transparent box, 6 of them offered justifications involving the transparency of the box and being able to encounter an object there (e.g., "probably just because it was clear so I could see something in there"). The remaining 3 participants could not explain their initial action. Amongst the 7 participants in the urgent condition who first reached for the opaque box, 5 of them offered justifications involving object types (e.g., "I knew that this [in transparent box] was closed and I didn't know this one [in opaque box] would be closed so I just opened it assuming it would be open"). The remaining 2 participants could not explain their initial action (for response examples see Supplementary Materials).

We also calculated the amount of time (in milliseconds) it took each participant to make a first reach to the edge of either box from the onset of the sound cue (videos at 30 frames per second were played back on a frame-by-frame basis). One participant's response was removed from analysis in the non-urgent condition as their latency was 3 standard deviations above the condition mean. A 2 (condition: urgent or non-urgent) x 2 (first box reach: transparent or opaque) between-subjects ANOVA was not performed as there were no first-reaches to the transparent box in the non-urgent condition. There was a main effect of condition on latency of first reach, $F(1, 29) = 5.42$, $p = 0.027$, $\eta_p^2 = 0.157$; participants in the urgent condition were faster to execute a first reach to a box ($M$ = 2941.67 milliseconds; $SD$ = 2038.32; $SEM$ = 509.58; $n$ = 16) than participants in the non-urgent condition ($M$ = 5848.89 milliseconds; $SD$ = 4535.20; $SEM$ = 1170.98; $n$ = 15). The main effect of first box reach on latency just reached significance, $F(1, 29) = 4.17$, $p = 0.05$, $\eta_p^2 = .126$. First reach was faster to the transparent box ($M$ = 2318.52 milliseconds; $SD$ = 1110.44; $SEM$ = 370.15; $n$ = 9) than to the opaque box ($M$ = 5178.79 milliseconds; $SD$ = 4103.37; $SEM$ = 874.84; $n$ = 22).

To summarize, our hypotheses were supported. First, in the non-urgent situation all first reaches were to the opaque box whereas in the urgent situation, many participants reached first to the transparent box. This suggests that adults tend to use rudimentary mindreading inferences in an urgent situation when an efficient decision has to be made about the types of object present. Second, latency analyses revealed that first reach in the urgent condition was faster than first reach in the non-urgent condition and, further, that first reach to the transparent box was faster than first reach to the opaque box. The response latencies fit with Apperly and Butterfill's (2009) broader theorizing that representing mental states as such is associated with deep and lasting demands on general cognitive resources in adults, and is incompatible with notions that human beings are (automatically) set to engage in abstract psychological reasoning.

**4. General Discussion**

The classical view of conceptual change in preschoolers' belief understanding is facing significant and increasing challenge from VOE-based studies showing that infants, in their looking-time expectations, are sensitive to others' false beliefs in a range of situations. The problem is that, whilst measuring indirect behavior has led to impressive advances in the TOM field (Baillargeon et al., 2010), looking-time data alone cannot definitively answer the question of whether indirect responses are driven by an understanding of belief as such, statistical learning, perceptual novelty, or perhaps some tracking of rudimentary belief-like states (Apperly & Butterfill, 2009; Heyes, 2014; Ruffman, 2014). Moreover, researchers are divided about which aspect of representational understanding certain VOE scenarios might be tapping into (e.g., false belief about object identity or false belief about object types – Scott & Baillargeon, 2009 versus Butterfill & Apperly, 2013). New insights into the development and cognitive processes of belief reasoning can be gained by going beyond infants' and preschoolers' task performances to consider the mature mindreading system that children grow into (Apperly, Samson, & Humphreys, 2009). In order to advance theoretical understanding of how VOE data should be considered, we sought to provide complementary evidence of how mature mindreaders would interpret the plot of VOE tasks.

Experiment 1 revealed that adults perceived the plot of Onishi and Baillargeon's (2005) VOE task to be meaningful. Adults spontaneously (when motivated with the belief instruction) provided

cognitively-oriented narratives to explain the unexpected outcome event sequence. Regardless of how adults narrated the event sequences, the majority of participants judged the unexpected outcome as being unexpected and the expected outcome as being expected. Adults were divided in their views of the event sequence adapted from Song and Baillargeon's (2008) VOE task. Participants clearly judged the expected outcome to be expected (referring to the agent's internal perceptions and cognitions over the tail box) but unexpected outcome judgments were mixed. These findings underscore Apperly's (2011) important point about how mental states have tenuous, holistically mediated relations to behavior and it can be challenging to constrain which aspects of an agent's situation may be relevant for interpreting his or her behavior.

Adults' narratives and judgments over Scott and Baillargeon's (2009) identity task are especially relevant in highlighting that researchers' own interpretations about where it makes sense for an agent to look are only obvious given a host of background or taken-for-granted assumptions. There were three key findings. First, this VOE scenario was the most challenging for participants to provide narrations for. Participants simply described the actions and did not make reference to internal cognitions, regardless of whether or not they were told to track beliefs. Second, participants judged the unexpected outcome as being expected, perceiving that it was reasonable for the agent to reach first for something that was being encountered rather than taking a chance on something that was not encountered. Third, even amongst participants who judged the expected outcome as being expected, the explanations revolved around the agent making a decision based on the types of objects present rather than on numerical identity per se. Our findings indicate that, at least from the mouths of adults, Scott and Baillargeon's VOE scenario taps reasoning about object types rather than identity (Butterfill & Apperly, 2013).

Experiment 2 reinforced the findings of Experiment 1: adults were more likely (and faster) to make a non-belief-based action and reach first for the transparent box where there was an object encountered, when the task was embedded in an urgent context as compared to a non-urgent context. If adults are not set to engage in abstract belief-based reasoning either when observing or participating in Scott and Baillargeon's VOE scenario, why might 18-month-olds look longer at the unexpected outcome compared to the expected outcome? One explanation could partly involve the fact that in the

familiarization trials, infants never saw the agent reach for or place a key inside the indivisible penguin. When the agent reached for the indivisible penguin at the test phase, infants' surprise may be due to the first occurrence of this novel agent-object relation (Heyes, 2014; Ruffman, 2014). Given that adults do ordinarily frame Scott and Baillargeon's VOE scenario in terms of sophisticated inferences involving misrepresentations over how a particular object is experienced, researchers should be cautious in drawing firm conclusions that infants are processing VOE scenarios in terms of abstract representations of mental states as such. On a positive note, our findings suggest that there is at least construct validity in the object location false belief VOE event sequence as compared to other more complex scenarios.

Our findings also connect with literature that speaks to the cognitive basis of human mindreading more generally. In Experiment 1, adults only spontaneously included inferences about others' mental states when they were instructed to consider other agents' beliefs. These results complement work by Apperly and colleagues' (2006) showing that false beliefs are not ascribed automatically. The case for non-automaticity of belief reasoning is also supported by the latency findings from Experiment 2 showing that incorrect first reach actions (actual location of the target) are faster than correct first reach actions (believed location of the target). If, as popularly assumed, adults (and infants and preschoolers) are from the start set to infer and put to use attributions about belief as such, correct responses should be faster (Perner & Roessler, 2012). It is reassuring that adults' latency data dovetailed with Atance et al.'s (2010) analysis of 3-year-olds' answer latencies on a battery of standard false belief tasks; the latency of young preschoolers' incorrect false belief answering were also faster than their correct latencies. Our findings, highlighting the non-automaticity of belief reasoning, when coupled with other evidence documenting that some mindreading situations are solved in a relatively automatic fashion (e.g., Van der Wel, Sebanz, & Knoblich, 2014), could suggest that mature mindreading involves multiple systems for tracking mental states with complementary trade-offs between efficiency and flexibility (Apperly & Butterfill, 2009; Low et al., 2016).

Researchers use the VOE paradigm to measure non-verbal (implicit) reasoning but Experiment 1 was limited to assaying participants' verbal (explicit) reasoning. Our approach is a necessary step towards breaking the interpretation deadlock facing VOE research. Adults' interpretations of the

meaning behind VOE event sequences and their judgments about what information may (or may not) be relevant to explaining action outcomes can partly help contextualize infants' understanding of the same stimuli, so we may better theorize about where development occurs and how. We addressed the limitation of Experiment 1 to some extent by measuring adults' non-verbal (reaching) responses in Experiment 2, and adults' actions dovetailed with adults' narratives about certain VOE event sequences. Nevertheless, future research could measure the amount of time adults spend looking at different VOE outcomes in false belief, true belief and ignorance conditions (see Silva, Ten Hope, & Tucker, 2014, for methodological suggestions), to gauge the degree of continuity between infants' and adults' mindreading ability. Examination of adults' looking responses on VOE tasks could even provide data on the extent to which adults track the agent's belief even when they were not instructed to do so (e.g., Schneider, Nott, & Dux, 2014). It will also be worthwhile to recruit a larger sample size to run our instruction manipulation as a within-subjects factor and, in so doing, profile individual differences in mindreading (e.g., Bukowski & Samson, 2017; Dodell-Feder, Lincoln, Coulson, & Hooker, 2013) that relate to when adults track or fail to track mental states in their verbal and non-verbal responses.

Our current findings offer up an important piece of practical advice: given partisan debates over how VOE tasks and data are to be interpreted, it would be theoretically informative for future VOE studies to compare and contrast infants' performance alongside adults' performance. Points of convergence and divergence between infants' and adults' reasoning can ultimately help specify the combination of cognitive systems, representations and processes that guide human mindreading.

**References**

Abell, F., Happé, F., & Frith, U. (2000). Do triangles play tricks? Attribution of mental states to animated shapes in normal and abnormal development. *Cognitive Development, 15*(1), 1–16. http://dx.doi.org/10.1016/S0885-2014(00)00014-9

Apperly, I. A. (2011). *Mindreaders: The Cognitive Basis of "Theory of Mind."* New York: Psychology Press.

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

Apperly, I. A., Riggs, K. J., Simpson, A., Chiavarino, C., & Samson, D. (2006). Is belief reasoning automatic? *Psychological Science, 17*(10), 841–844. https://dx.doi.org/10.1111/j.1467-9280.2006.01791.x

Apperly, I. A., Samson, D., & Humphreys, G. W. (2009). Studies of adults can inform accounts of theory of mind development. *Developmental Psychology*, *45*(1), 190–201. https://dx.doi.org/10.1037/a0014098

Atance, C. M., Bernstein, D. M., Meltzoff, A. N. (2010). Thinking about false belief: It's not just what children say but how long it takes them to say it. *Cognition, 116*(2), 297–301. http://dx.doi.org/10.1016/j.cognition.2010.05.008

Baillargeon, R., Scott, R. M., & He, Z. (2010). False-belief understanding in infants. *Trends in Cognitive Sciences, 14*(3), 110–117. http://dx.doi.org/10.1016/j.tics.2009.12.006

Butterfill, S. A. & Apperly, I. A. (2013). How to construct a minimal theory of mind. *Mind & Language, 28*(5), 606–637. http://dx.doi.org/10.1111/mila.12036

Bukowski, H., & Samson, D. (2017). New insights into the inter-individual variability in perspective taking. *Vision, 1*(1), 8. http://dx.doi:10.3390/vision1010008

Carlson, S. M., Claxton, L. J., & Moses, L. J. (2015). The relation between executive function and theory of mind is more than skin deep. *Journal of Cognition and Development, 16*(1), 186–197. http://dx.doi.org/10.1080/15248372.2013.824883

Dodell-Feder, D., Lincoln, S. H., Coulson, J. P., & Hooker, C. I. (2013). Using fiction to assess mental state understanding: A new task for assessing theory of mind in adults. *PLoS One, 8*(11), e81279. https://dx.doi.org/10.1371/journal.pone.0081279

Edwards, K., & Low, J. (2016). Reaction time profiles of adults' action prediction reveal two mindreading systems. *Cognition*, 160(1), 1–16. http://dx.doi.org/10.1016/j.cognition.2016.12.004

Fizke, E., Butterfill, S. A., & Rakoczy, H. (2013). *Toddlers' understanding of false belief about an object's identity*. Poster presented at the Meeting of the Society for Research in Child Development. Seattle, USA.

Haith, M. M. (1998). Who put the cog in infant cognition? Is rich interpretation too costly? *Infant Behavior and Development, 21*(2), 167–179. http://dx.doi.org/10.1016/S0163-6383(98)90001-7

Hespos, S., Gredebäck, G., von Hofsten, C., & Spelke, E. S. (2009). Occlusion is hard: Comparing predictive reaching for visible and hidden objects in infants and adults. Cognitive Science, 33(8), 1484–1502. http://dx.doi.org/10.1111/j.1551-6709.2009.01051.x

Heyes, C. (2014). False belief in infancy: A fresh look. *Developmental Science, 17*(5), 1–13. http://dx.doi.org/10.1111/desc.12148

Leavitt, J. D., & Christenfeld, N. J. (2011). Story spoilers don't spoil stories. *Psychological Science, 22*(9), 1152–1154. http://dx.doi.org/10.1177/0956797611417007

Lovett, S. B., & Pillow, B. H. (2010). Age-related changes in children's and adults' explanations of interpersonal actions. *The Journal of Genetic Psychology, 171*(2), 139–167. http://dx.doi.org/10.1080/00221320903548100

Lagattuta, K. H. (2005). When you shouldn't do what you want to do: Young children's understanding of desires, rules, and emotions. *Child Development, 76*(3), 713–733. http://dx.doi.org/10.1111/j.1467-8624.2005.00873.x

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

Low, J., & Perner, J. (2012). Implicit and explicit theory of mind: State of the art. *British Journal of Developmental Psychology, 30*(1), 1–13. http://dx.doi.org/10.1111/j.2044-835X.2011.02074.x

Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? Science, 308(5719), 255–258. http://dx.doi.org/10.1126/science.1107621

Perner, J. (1991). *Understanding the Representational Mind*. Cambridge, MA: Bradford Books / MIT Press.

Perner, J., & Roessler, J. (2012). From infants' to children's appreciation of belief. *Trends in Cognitive Sciences, 16*(10), 519–525. http://dx.doi.org/10.1016/j.tics.2012.08.004.

Rakoczy, H., Bergfeld, D., Schwarz, I., & Fizke, E. (2015). Explicit theory of mind is even more unified than previously assumed: Belief ascription and understanding aspectuality emerge together in development. *Child Development, 86*(2), 486–502. http://doi.org/10.1111/cdev.12311

Ruffman, T. (2014). To belief or not belief: Children's theory of mind. *Developmental Review, 34*(3), 265–293. http://doi.org/10.1016/j.dr.2014.04.001

Sabbagh, M. A., Xu, F., Carlson, S. M., Moses, L. J., & Lee, K. (2006). The development of executive functioning and theory-of-mind: A comparison of Chinese and U.S. preschoolers. *Psychological Science, 17*(1), 74–81. http://dx.doi.org/10.1111/j.1467-9280.2005.01667.x

Schneider, D., Nott, Z. E., & Dux, P. E. (2014). Task instructions and implicit theory of mind. *Cognition*, *133*(1), 43–47. http://dx.doi.org/10.1016/j.cognition.2014.05.016

Scott, R. M., & Baillargeon, R. (2009). Which penguin is this? Attributing false beliefs about object identity at 18 months. *Child Development, 80*(4), 1172–119. http://dx.doi.org/10.1111/j.1467-8624.2009.01324.x

Silva, F. J., Ten Hope, M. I., & Tucker, A. L. (2014). Adult humans' understanding of support relations: An up-linkage replication. *Learning & Behavior*, *42*(4), 337–347. http://dx.doi.org/10.3758/s13420-014-0151-0

Sodian, B. (2011). Theory of mind in infancy. *Child Development Perspectives, 5*(1), 39–43. http://dx.doi.org/10.1111/j.1750-8606.2010.00152.x

Sodian, B. (2016). Is false belief understanding continuous from infancy to preschool age? In D. Barner & A. S. Baron (Eds.), *Core Knowledge and Conceptual Change* (p. 301–320). New York: Oxford University Press.

Song, H. J., & Baillargeon, R. (2008). Infants' reasoning about others' false perceptions. *Developmental Psychology, 44*(6), 1789–1795. http://dx.doi.org/10.1037/a0013774

Shapiro, L. R., & Hudson, J. A. (1991). Tell me a make-believe story: Coherence and cohesion in young children's picture-elicited narratives. *Developmental Psychology, 27*(6), 960–974. http://dx.doi.org/10.1037/0012-1649.27.6.960

Van der Wel, R.P.R.D., Sebanz, N., & Knoblich, G. (2014). Do people automatically track others' beliefs? Evidence from a continuous measure. *Cognition, 130*(1), 128–133. http://dx.doi.org/10.1016/j.cognition.2013.10.004

Wellman, H. M. (2014). *Making minds*: *How theory of mind develops.* New York: Oxford University Press.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development, 72*(3), 655–684. http://dx.doi.org/10.1111/1467-8624.00304

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13*(1), 103–128. http://doi.org/10.1016/0010-0277(83)90004-5

Yott, J., & Poulin-DuBois, D. (2016). Are infants' theory-of-mind abilities well integrated? Implicit understanding of intentions, desires, and beliefs. *Journal of Cognition and Development, 17*(5) 683–698. http://dx.doi.org/10.1080/15248372.2015.1086771

Zawidzki, T. W. (2013). Mindshaping: A new Framework for Understanding Human Social Cognition. Cambridge, MA: MIT Press.

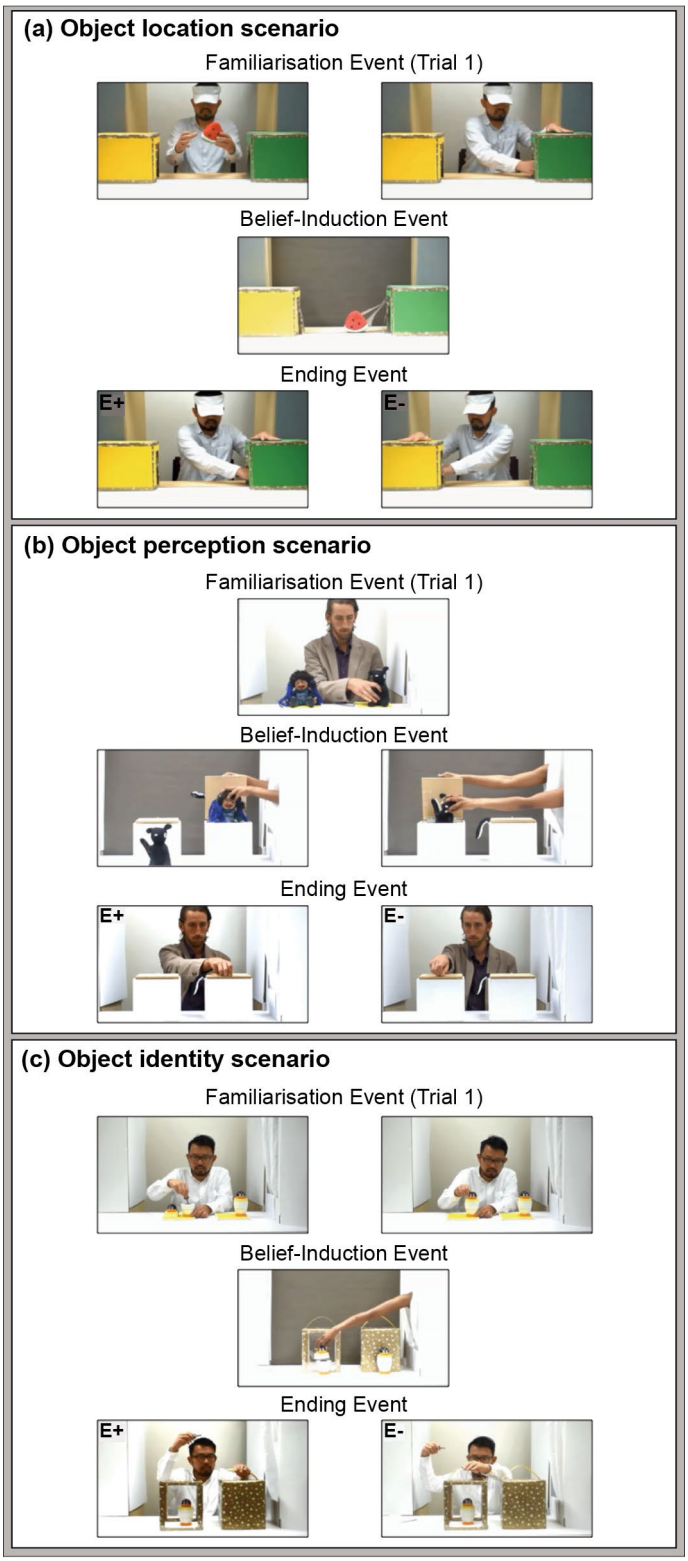


**Fig. 1.** Film stills from the (a) object location, (b) object perception and (c) object identity false belief scenarios, showing the familiarization events, the belief-induction events and the ending events (where the outcome is either expected E+ or unexpected E-).

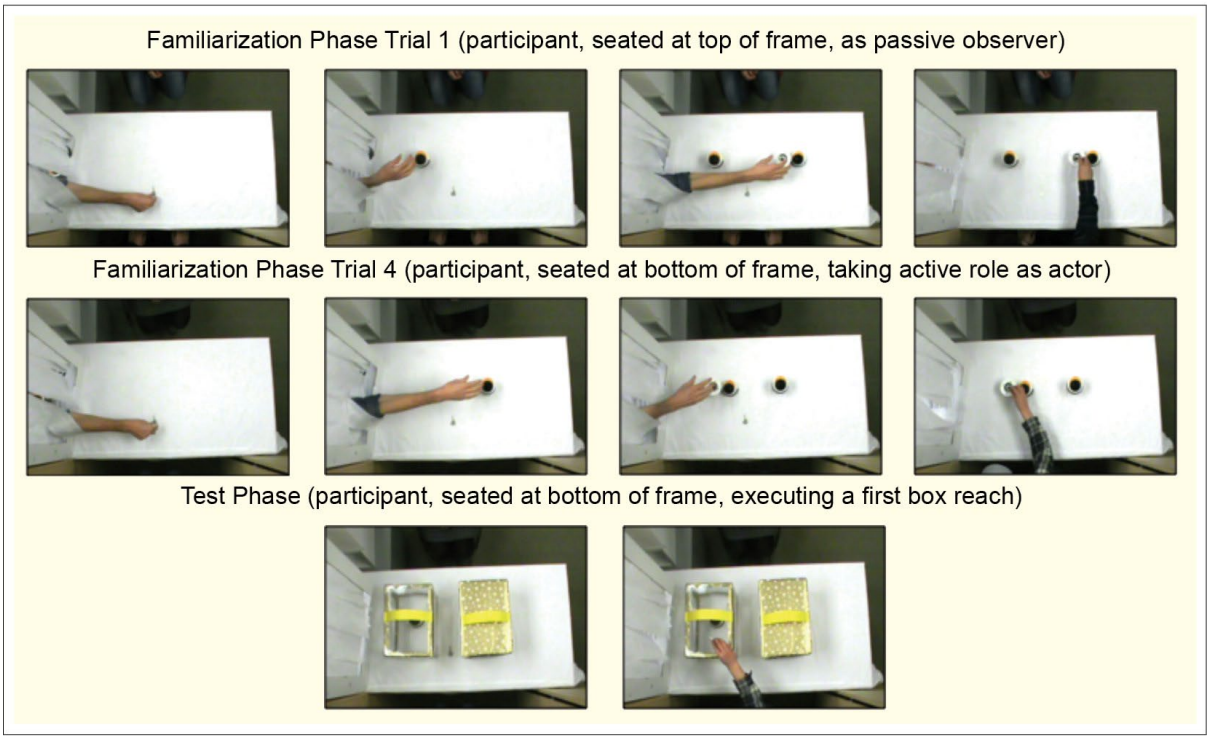


**Fig. 2.** Film stills showing the familiarization and test phases of the identity scenario in Experiment 2. In familiarization trial 1, the participant (seated at top) passively watched the sequence of events: the hand placed the key on the stage, followed by the indivisible toy and the divisible toy, and then the actor-agent placed the key into the divisible toy and assembled it to deactivate the siren alarm. In familiarization trial 4, the participant (seated at bottom) actively played the role of the actor-agent, and placed the key inside the divisible toy and assembled it to deactivate the siren. In the test phase, the blind was pulled up and the siren was immediately triggered whilst the participant (seated at bottom) was confronted with a transparent box (containing the assembled divisible toy) and an opaque box (containing the indivisible toy). In this example, the participant first reached to the transparent box.

**Table 1.** Mean narrative context scores (with standard deviation and standard error of mean) for the different VOE scenarios by instruction and outcome shown.

| Scenario | Instruction | Outcome | Mean (SD; SEM) | |
|---|---|---|---|---|
| Location | Neutral | Unexpected | 2.33 (0.78; 0.22) | |
| | | Expected | 2.25 (0.45; 0.13) | |
| | Object | Unexpected | 1.92 (0.29; 0.08) | |
| | | Expected | 2.17 (0.39; 0.11) | |
| | Belief | Unexpected | 3.08 (0.67; 0.19) | * |
| | | Expected | 2.42 (0.67; 0.19) | |
| Perception | Neutral | Unexpected | 2.08 (0.29; 0.08) | |
| | | Expected | 2.58 (0.90; 0.26) | |
| | Object | Unexpected | 1.83 (0.39; 0.11) | * |
| | | Expected | 2.42 (0.67; 0.19) | |
| | Belief | Unexpected | 3.17 (0.83; 0.24) | |
| | | Expected | 2.67 (0.78; 0.22) | |
| Identity | Neutral | Unexpected | 2.08 (0.29; 0.08) | |
| | | Expected | 2.25 (0.62; 0.18) | |
| | Object | Unexpected | 1.83 (0.39; 0.11) | |
| | | Expected | 1.92 (0.29; 0.08) | |
| | Belief | Unexpected | 2.50 (0.80; 0.23) | |
| | | Expected | 2.67 (0.89; 0.26) | |

Note: *$p$ = 0.02