CHILD DEVELOPMENT

Child Development, xxxx 2017, Volume 00, Number 0, Pages 1–14

Inferring Beliefs and Desires From Emotional Reactions to Anticipated and Observed Events

Yang Wu and Laura E. Schulz
MIT

Researchers have long been interested in the relation between emotion understanding and theory of mind. This study investigates a cue to mental states that has rarely been investigated: the dynamics of valenced emotional expressions. When the valence of a character's facial expression was stable between an expected and observed outcome, children (N = 122; M = 5.0 years) recovered the character's desires but did not consistently recover her beliefs. When the valence changed, older but not younger children recovered both the characters' beliefs and desires. In contrast, adults jointly recovered agents' beliefs and desires in all conditions. These results suggest that the ability to infer mental states from the dynamics of emotional expressions develops gradually through early and middle childhood.

Researchers have long noted correlations between the development of children's belief–desire psychology and their understanding of emotions (e.g., Bartsch & Estes, 1996; De Rosnay, Fink, Begeer, Slaughter, & Peterson, 2014; Harwood & Farrar, 2006; Hughes & Dunn, 1998; LaBounty, Wellman, Olson, Lagattuta, & Liu, 2008; Wellman, 2014; Widen, 2013; Widen & Russell, 2008), and proposed that children construct an intuitive theory of mind in which beliefs, desires, and emotions are causally linked (e.g., Harris, Johnson, Hutton, Andrews, & Cooke, 1989; Thompson & Lagattuta, 2006; Lagattuta, Wellman, & Flavell, 1997; Lagattuta & Wellman, 2001; see Harris, 2008 and Wellman, 2014 for discussion and review). Here we ask to what extent this intuitive theory allows children to recover others' beliefs and desires from their emotional reactions to events.

Given the extensive history of work on theory of mind and emotion, some justification is required for posing this as an unanswered question. Note, however, that much of this research has looked at children's ability to use knowledge of others' beliefs and desires to predict their emotions. Here we are interested in the inverse problem: children's ability to use emotional expressions to recover other mental states. To follow, we briefly summarize past research and then ask whether children can use someone's emotional reactions to anticipating and observing an outcome to jointly infer her beliefs and desires.

Human beings are sensitive to others' emotional expressions from birth (Field, Woodson, Greenberg, & Cohen, 1982). Infants show different patterns of behavior in response to happy, fearful, sad, and angry faces and voices (Field et al., 1982; Haviland & Lelwica, 1987; Montague & Walker-Andrews, 2001); represent emotions cross-modally (Walker-Andrews, 1997); and discriminate expressions even within valence (e.g., matching happy faces to happy voices and interested faces to interested voices; Soken & Pick, 1999; see also Flom & Bahrick, 2007; Haviland & Lelwica, 1987; Hoehl & Striano, 2008; Soderstrom, Reimchen, Sauter, & Morgan, 2017). Older infants check their parents' faces given ambiguous stimuli and approach or retreat depending on the valence of the parents' facial expression (e.g., Hornik & Gunnar, 1988; Moses, Baldwin, Rosicky, & Tidball, 2001; Mumme & Fernald, 2003; Sorce, Emde, Campos, & Klinnert, 1985; Walden & Ogan, 1988).

This material is based on the work supported by the Center for Brains, Minds and Machines (CBMM), funded by NSF STC Award CCF-1231216. We thank the Boston Children's Museum and participating parents and children. We also thank Emily Tsang, Veronika Jedryka, Bonnie Wong, Margaret Lattanzi-Silveus, Kyla Truman, and Michelle Wang for help with data collection.

Correspondence concerning this article should be addressed to Yang Wu, Department of Brain and Cognitive Sciences, MIT, 77 Massachusetts Avenue, Cambridge, MA 02139. Electronic mail may be sent to yangwu@mit.edu.

0009-3920/2017/xxxx-xxxx
DOI: 10.1111/cdev.12759

Infants might respond differentially to emotional expressions without understanding emotions per se; however, recent research suggests that even infants relate emotional expressions to goal-directed actions. As early as 8 months, infants expect agents to express positive rather than negative emotions when they achieve their goals (Skerry & Spelke, 2014). By age 2, children explicitly predict that an agent will be happy if her desires are fulfilled and sad if they are thwarted (e.g., Wellman & Woolley, 1990; Yuill, 1984). In 2-year-olds, these emotion concepts may be relatively undifferentiated, distinguishing primarily between positive and negative valences (see, e.g., Widen & Russell, 2008, 2010); nonetheless, children appropriately map goal fulfillment onto positive emotions and goal failure to negative ones.

More mixed findings have been obtained for belief inferences. As young as 3, children seem to experience suspenseful emotions in response to false belief scenarios; they are more likely to furrow their brow and bite their lips when they see that someone is about to act consistent with a false (vs. true) belief (Moll, Kane, & McGowan, 2016). Between 4 and 6, children become increasingly likely to predict that someone else will feel surprised if her beliefs are violated and feel happy if she falsely believes that an action will fulfill her desires (Hadwin & Perner, 1991; Harris et al., 1989; Wellman & Banerjee, 1991). However, considerable research suggests that children's ability to attribute the emotional reactions generated by true and false beliefs lags behind their ability to explicitly represent the beliefs themselves (e.g., De Rosnay, Pons, Harris, & Morrell, 2004; Hadwin & Perner, 1991; Harris et al., 1989; Pons, Harris, & De Rosnay, 2004; Ruffman & Keenan, 1996; Wellman & Bartsch, 1988). Thus, 4- and 5-year-olds may know that Red Riding Hood falsely believes her grandmother is in bed and nonetheless conclude that Red Riding Hood is frightened (Bradmetz & Schneider, 1999).

Fewer studies have looked at children's ability to reason backward from emotional reactions to desires and beliefs, and again the strongest evidence is for inferences about desires. By 18 months, infants can use an agent's verbal cues ("Yummy!" vs. "Yucky!") together with her emotional expressions to decide if she wants a food different from what the child herself wants (e.g., broccoli rather than goldfish crackers; Repacholi & Gopnik, 1997). Similarly, 2- and 3-year-olds can use someone's emotional reaction to infer whether she is looking at desirable crackers or undesirable broccoli (Wellman, Phillips, & Rodriguez, 2000). By preschool, children map happy and sad emotional reactions onto familiar desirable or undesirable events (e.g., getting a puppy or dropping an ice cream cone; Denham, Zoller, & Couchoud, 1994; Fabes, Eisenberg, McCormick, & Wilson, 1988; Gnepp, McKee, & Domanic, 1987; Harris, Olthof, Terwogt, & Hardman, 1987; Widen & Russell, 2010).

However, children younger than 6 rarely refer to agents' beliefs in explaining others' emotional reactions (Rieffe, Terwogt, & Cowan, 2005). One exception is that 4-year-olds mention beliefs in explaining fearful reactions (e.g., saying "she thought it was a ghost" if a character is scared at hearing a noise) and atypical ones (saying "she thought it would be something else" if a character is sad upon opening a present; Rieffe et al., 2005). Four-year-olds also spontaneously refer to beliefs in explaining surprise or curiosity and do so more often given these "epistemic" emotions than happy or sad ones (Wellman & Banerjee, 1991).

As noted earlier, however, by preschool, children have learned scripts relating familiar events and emotions (e.g., Fabes et al., 1988; Gnepp et al., 1987; Harris et al., 1987; Widen & Russell, 2010). Children might link fear and ghosts or a disappointing gift with sadness (Rieffe et al., 2005) without necessarily reasoning about the relation between emotions and beliefs more broadly. Similarly, children might guess that an agent did not know about or expect unusual or mysterious events ("she didn't think there would be a giraffe"; "she didn't know what was in the box"; Wellman & Banerjee, 1991), because the events are atypical or mysterious rather than because they reason about the beliefs underlying emotional reactions generally.

Perhaps the strongest support for the idea that children infer the thoughts underlying emotional responses comes from work showing that children invoke others' thoughts about the past to explain their emotions in the present (e.g., Harris, Guz, Lipian, & Man-Shu, 1985; Lagattuta et al., 1997; Lagattuta & Wellman, 2001; Taylor & Harris, 1983). Four-year-olds recognize that people respond more intensely to recent events than past events; by 6, children recognize that people's responses to events depend on whether they remember them and that people are happier remembering positive experiences and forgetting negative ones (e.g., Harris et al., 1985; Taylor & Harris, 1983). Children also understand that individuals' particular histories can lead to idiosyncratic emotional responses. If a girl's doll is broken by a clown, children predict that she will be sad on seeing another clown and explain her sadness by saying she is thinking about her doll (Lagattuta & Wellman, 2001; Lagattuta et al., 1997).

Findings like these suggest that children understand that thoughts affect feelings. However, they do not address the question of whether children can use someone's emotional reactions to recover the content of her otherwise unknown thoughts. Theory of mind is a hard inference problem because it often involves situations where the only information available is that which can be gleaned from the environment and observed behavior. In these cases, an observer has no more access to others' past history of emotional experiences than she does to their beliefs and desires. However, it is precisely in such contexts that others' emotional reactions might be an especially valuable cue to their mental states.

Collectively, these findings leave open the question of the degree to which children can use their understanding of emotions to gain insight into others' minds. It may be difficult or impossible (in the absence of extensive prior knowledge) to recover the representations driving the changes in emotional expressions when these changes are unrelated to external, observable events (e.g., as when someone looks happy at one moment and sad the next simply because they first think about a joyous event and then an unhappy one). However, it may be possible to recover others' mental states when these changes are probabilistically associated with external events. Here we focus on agent's emotional expressions in response to testimony about, and observation of, an event. When an agent's mental states are otherwise underdetermined by her actions and the context, can children compare her emotional reactions to an expected and observed outcome to jointly recover her desires and beliefs?

Previous work on inferring beliefs from emotions has found both successes and failures between ages 4 and 6 (e.g., Hadwin & Perner, 1991; Harris et al., 1989; Rieffe et al., 2005; Wellman & Banerjee, 1991), thus we focus our investigation on the same age range. In contrast to most previous work, we focus (a) specifically on "backwards inferences," from emotional reactions to beliefs and desires; (b) on children's ability to recover both beliefs and desires simultaneously; (c) on inferences about ordinary (rather than unusual, surprising) events; and (d) on a cue to agents' mental states that has rarely been investigated: the dynamics of valenced facial expressions. We predict that children may be able to use an agent's emotional reactions to anticipated and observed outcomes to recover both her beliefs and desires. Specifically, given the emotional reaction to the observed outcome, children may be able to infer the agent's desires; given the presence or absence of a change in valence between anticipated and observed outcomes, children may be able to infer the content of the agent's initial beliefs.

Finally, because the question of interest here is whether children can infer mental states from the dynamics of emotional reactions to expected and observed outcomes, we will elide a question that has been the focus of many previous investigations: How can children draw inferences from emotional reactions to the emotions themselves? (e.g., Ekman & Oster, 1979; Gross & Ballif, 1991; Izard, 1994; Widen & Russell, 2008, 2010). We take as a premise that, within a well-specified context and shared cultural knowledge, 4- and 5-year-olds can use prototypical happy and sad facial expressions to infer happiness and sadness (see, e.g., Widen, 2013). Our question is whether children can use information in others' emotional responses to anticipated and observed outcomes to jointly infer their beliefs and desires.

Experiment 1

Method

Participants

Thirty-two children (M = 5.0 years, range: 4.1–5.9; 50% girls) were recruited from an urban children's museum between April and July 2014. Although most of the children were White and middle class, a range of ethnicities and socioeconomic backgrounds reflecting the diversity of the local population (47% European American, 24% African American, 9% Asian, 17% Latino, 4% two or more races) and the museum population (29% of museum attendees receive free or discounted admission) were represented throughout.

To follow and throughout, we treat age as a continuous variable and then perform post hoc analyses in age bins (4-year-olds and 5-year-olds) in order to enable comparison with the previous literature on theory of mind, which has largely treated age groups categorically (see Wellman, 2014 for review). To ensure a balanced distribution across ages, children were recruited in age bins consisting of 16 four-year-olds (M = 4.4 years, range: 4.1–4.8; 56% girls) and 16 five-year-olds (M = 5.5 years, range: 5.0–5.9; 44% girls).

Materials

Each child saw four illustrated stories (see Figure 1), two presenting *valence stable* conditions (happy–happy and sad–sad) and two presenting *valence change* conditions (happy–sad and sad–happy).

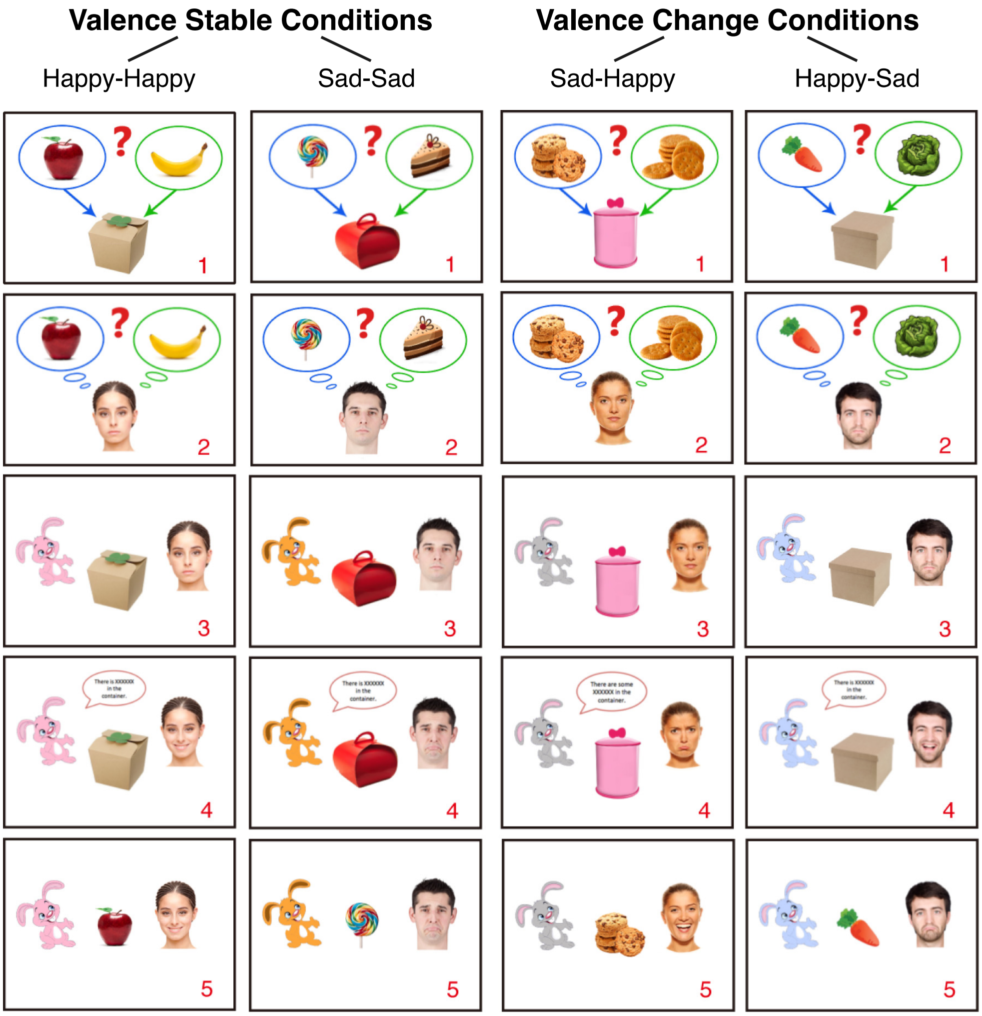

Figure 1. Examples of picture stimuli used in the happy–happy, sad–sad, sad–happy, and happy–sad conditions, respectively. The mapping between stories and conditions, the order of conditions, and the expected–actual contents of the containers were counterbalanced across participants. See text for details.

Because we used canonical happy and sad faces, and research suggests that by age 4, children can interpret these as such (see Widen, 2013 for review), expressions coded by the Facial Action Coding System (Ekman & Friesen, 1976) were not critical here; the facial expressions used here were from istock photos (http://www.istockphoto.com/). The mapping between stories and conditions, the order of conditions, and the expected–actual contents of the containers were counterbalanced across participants, resulting in a total of 16 storybooks. A different agent and a different bunny (indicated by its color) were used in each story.

Procedure

Children were tested individually; all sessions were videotaped. Children were asked check questions to encourage them to follow along. Incorrect responses were corrected throughout. (Collapsing data from all three experiments in the study, children's accuracies on the six check questions were .74, .73, .80, .99, .99, and 1.00, respectively. Incorrect responses consisted primarily of guessing: e.g., when asked "Do we know what is in the container?" or "Do we know what Sally wants to eat?," children who answered incorrectly would say "an apple" or "a banana" rather than say "no." No children were excluded on the bases of incorrect responses; however, none of the results change if only children who answered all six check questions correctly are included.)

Each story was read consecutively, as follows (using the apple–banana story as an example). The experimenter placed Picture 1 on the table and said, "This is a container. Sometimes there is an apple inside and sometimes there is a banana. But before we open it, we don't know what's inside." She introduced Picture 2 and said, "This is Sally. She wants something to eat. She might want an apple, or she might want a banana; but she hasn't told us." Children were asked (Check Questions 1 and 2): "Do we know what's in the container?" and "Do we know what Sally wants to eat?" The experimenter introduced Picture 3 and said, "This is the pink bunny. The bunny wants to help. So the bunny says: 'Hi, Sally! Let me tell you what's in the container!' But, the bunny could be right or wrong." Children were asked (Check Question 3): "Is the bunny always right?" The experimenter introduced Picture 4, with the facial expression appropriate to the condition and said, "The bunny tells Sally what he thinks is in the container. But he doesn't tell us. It's a secret. After hearing the bunny's secret, Sally's response is this . . ." Children were asked (Check Question 4): "Is she happy or sad?" The experimenter introduced Picture 5 and said, "Then the bunny opens the container and takes out what's inside." Children were asked (Check Question 5): "What's inside?" Pointing to Picture 5, the experimenter said, "After seeing what's actually in the container, Sally's response is this . . ." Children were asked (Check Question 6): "Is she happy or sad?"

Finally, children were asked two test questions, in fixed order. The experimenter pointed to Picture 5 and asked (desire question): "Based on this response, what did Sally want to eat today, an apple or a banana?" Then, she pointed to Picture 4 and asked (belief question): "Before the container was opened, but after hearing the bunny's secret, what did Sally think was inside: an apple or a banana?"

Coding

The first author (YW) coded all the responses to the two test questions offline from videotape. Seventy-five percent of these responses were recoded by an independent coder blind to hypotheses and conditions; there was 100% agreement on children's responses. Children's responses to the desire question were coded as "actual content" if they chose the content in the container and "alternative" if they chose the other food. Children's responses to the belief question were coded as "true belief" if they chose the content in the container and "false belief" if they chose the other food. Two of 256 responses could not be classified (i.e., "both" and "she did not know"); these were coded as missing values. (Categorizing these as wrong responses instead of missing values does not change the results here, or in the following experiments.)

Results and Discussion

As discussed (see Participants), we first analyze the effect of age as a continuous variable and follow-up with post hoc analyses by age group to enable comparison with the previous literature. Generalized estimating equations were used for all analyses except as indicated. Generalized estimating equations are comparable to repeated measures analysis of variance (ANOVA), but unlike repeated measures ANOVA, they are appropriate for categorical outcomes (as in the forced choice binary responses here) and robust to missing cells (see Coding). See Figure 2 for results.

Desires

There was a main effect of age, $b = -3.07$, $SE = 1.07$; $\chi^2(1, N = 128) = 8.20$, $p = .004$, and of the valence of the second facial expression, $b = 33.49$, $SE = 4.88$; $\chi^2(1, N = 128) = 12{,}913.87$, $p < .001$, on children's responses; the interaction was also significant, $b = 3.07$, $SE = 1.09$; $\chi^2(1, N = 128) = 7.92$, $p = .005$. Five-year-olds performed at ceiling; all children inferred that the agent wanted the food in the container when the final expression was happy (happy–happy, sad–happy) and the alternative when the expression was negative (sad–sad, happy–sad). Four-year-olds performed at ceiling when the expression was positive but had slightly more difficulty when the expression was negative (sad–sad: 12/16, $p = .077$, 95% CI [0.48, 0.93]; happy–sad: 13/16, $p = .021$, 95% CI [0.54, 0.96]; comparisons to chance by binomial test throughout).

Beliefs

There was no main effect of age, $b = 0.60$, $SE = 0.42$; $\chi^2(1, N = 126) = 1.09$, $p = .297$, but there was a main effect of the presence or absence of a valence change, $b = 9.67$, $SE = 3.62$; $\chi^2(1, N = 126) = 11.13$, $p < .001$, and an interaction, $b = -2.24$, $SE = 0.76$; $\chi^2(1, N = 126) = 8.82$, $p = .003$.

For 5-year-olds, the presence or absence of a valence change affected their responses, $b = -2.84$, $SE = 0.67$; $\chi^2(1, N = 63) = 18.20$, $p < .001$. Five-year-olds inferred false beliefs in the valence change conditions (sad–happy: 15/16, $p < .001$, 95% CI [0.70, 1.00]; happy–sad: 13/16, $p = .021$, 95% CI [0.54,

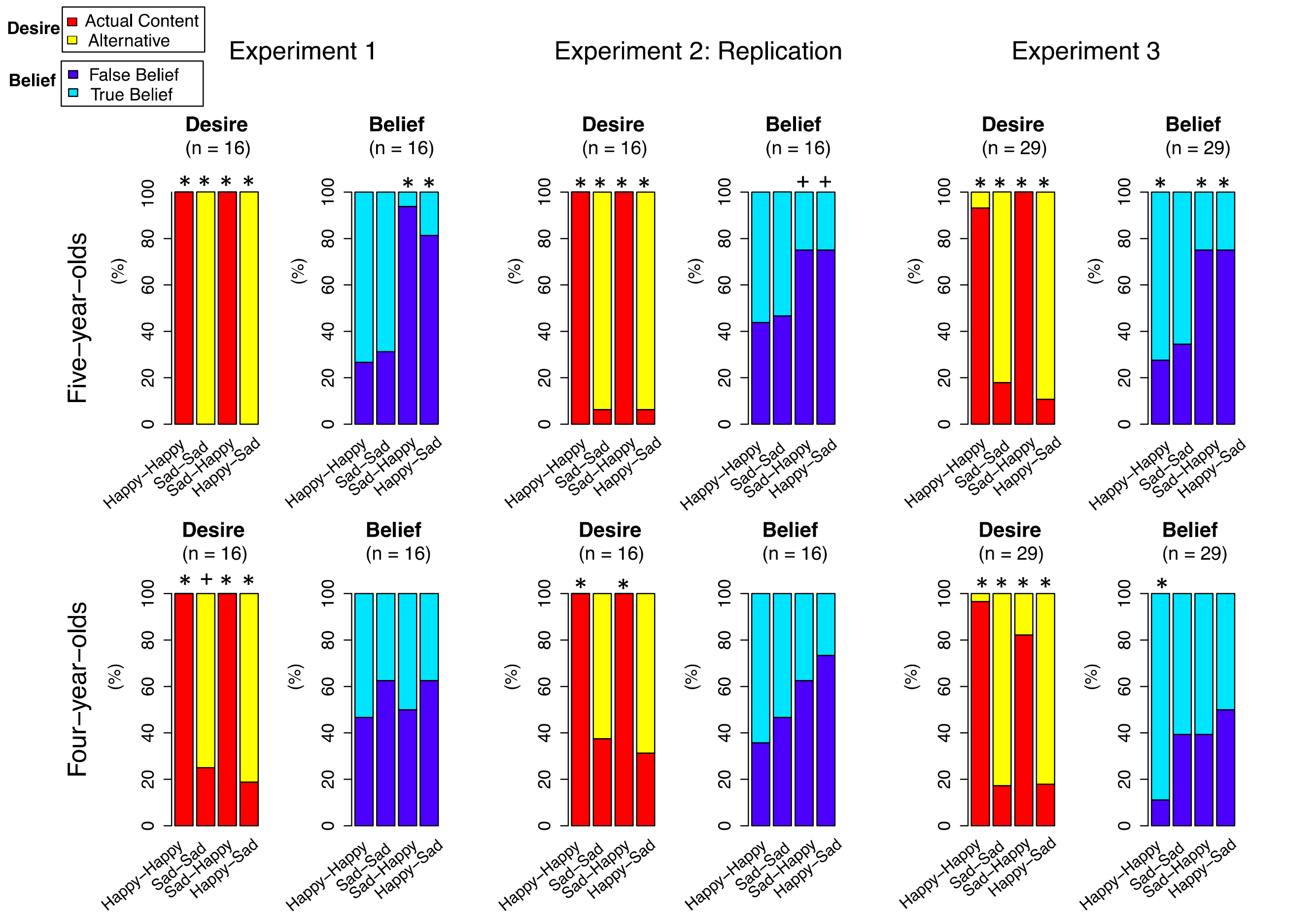

Figure 2. Five- and 4-year-olds' inferences about desires and beliefs in Experiments 1–3. $^{*}p < .05$. $^{+}p < .10$.

0.96]), but did not infer true beliefs in the valence stable conditions (happy–happy: 11/15, $p = .119$, 95% CI [0.45, 0.92]; sad–sad: 11/16, $p = .210$, 95% CI [0.41, 0.89]).

By contrast, 4-year-olds had difficulty recovering the agent's beliefs in both conditions. The effect of the presence or absence of a valence change was not significant, $b = -0.06$, $SE = 0.51$; $\chi^2(1, N = 63) = 0.01$, $p = .910$. Children performed at chance in all conditions (valence change, false belief responses: sad–happy: 8/16, $p = 1.000$, 95% CI [0.25, 0.75]; happy–sad: 10/16, $p = .455$, 95% CI [0.35, 0.85]; valence stable, true belief responses: happy–happy: 8/15, $p = 1.000$, 95% CI [0.27, 0.79]; sad–sad: 6/16, $p = .454$, 95% CI [0.15, 0.65]).

Experiment 2: Replication

Given concern about the reproducibility of scientific results (Open Science Collaboration, 2015), we conducted a preregistered replication (Open Science Framework: https://osf.io/e36vm/?view_only=643ddd21a06a49ca987105b1069b487a). The preregistered analyses were the generalized estimated equations by age group, predicting that the presence or absence of a valence change would affect belief inferences in 5- but not 4-year-olds (https://osf.io/kx97a/?view_only=5590ef3ed01548b4b963ba83667fdd63). We also tested adults to clarify children's responses to the valence stable conditions.

Method

Participants

Thirty-two children ($M = 4.9$ years, range: 4.0–5.9; 41% girls) were recruited from the museum between August and September 2015. To ensure a balanced distribution across ages, children were recruited in age bins consisting of 16 four-year-olds ($M = 4.4$ years, range: 4.0–4.9; 31% girls) and 16 five-year-olds (M = 5.4 years, range: 5.0–5.9; 50% girls).

Sixty-five adults were recruited on Amazon Mechanical Turk. Participation was restricted to individuals with IP addresses from the United States and with HIT approval rate of 95% or higher. Because Amazon Mechanical Turk workers are paid by the task, there is an incentive to rush through tasks. To ensure that online participants were attentive, check questions for the adults were used as exclusion criteria. Four participants were excluded for answering one or more check questions incorrectly (see Materials and Procedure).

Materials and Procedure

The materials and procedure for the children were identical to the initial experiment. Adults were tested online using the same materials except that all materials were written and only three of the check questions were asked: "Is the bunny always right?," "Is Sally happy or sad?" when Sally responded to the bunny's secret, and "Is Sally happy or sad?" when Sally responded to the actual contents in the container.

Coding

Coding was identical to the initial experiment. Seventy-five percent of responses were recoded by an independent coder blind to hypotheses and conditions; there was 99% agreement on children's responses. Four of 256 responses could not be classified (e.g., "lollipop and cake" and "I don't know"); as in Experiment 1, these responses were coded as missing values.

Results and Discussion

Desires

For children, there was a main effect of age, $b = -2.16$, $SE = 0.70$; $\chi^2(1, N = 128) = 9.70$, $p = .002$, and the valence of the final expression, $b = 36.85$, $SE = 3.28$; $\chi^2(1, N = 128) = 19{,}652.07$, $p < .001$; the interaction was also significant, $b = 2.16$, $SE = 0.72$; $\chi^2(1, N = 128) = 8.93$, $p = .003$. Five-year-olds performed near ceiling (happy–happy: 16/16, $p < .001$, 95% CI [0.97, 1.00]; sad–happy: 16/16, $p < .001$, 95% CI [0.97, 1.00]; sad–sad: 15/16, $p < .001$, 95% CI [0.70, 1.00]; happy–sad: 15/16, $p < .001$, 95% CI [0.70, 1.00]). Four-year-olds performed at ceiling when the final expression was positive but again had more difficulty when the expression was negative (happy–happy: 16/16, $p < .001$, 95% CI [0.97, 1.00]; sad–happy: 16/16, $p < .001$, 95% CI [0.97, 1.00]; sad–sad: 10/16, $p = .455$, 95% CI [0.35, 0.85]; happy–sad: 11/16, $p = .210$, 95% CI [0.41, 0.89]).

Beliefs

There was a main effect of the presence or absence of a valence change, $b = 1.81$, $SE = 3.22$; $\chi^2(1, N = 124) = 8.95$, $p < .003$. Neither the main effect of age, $b = 0.26$, $SE = 0.42$; $\chi^2(1, N = 124) = 0.01$, $p = .941$, nor an interaction between age and a valence change, $b = -0.60$, $SE = 0.66$; $\chi^2(1, N = 124) = 0.84$, $p = .360$) was significant.

Again, 5-year-olds' inferences were affected by the presence or absence of a valence change, $b = -1.29$, $SE = 0.55$; $\chi^2(1, N = 63) = 5.63$, $p = .018$. In the valence change conditions, children tended to infer false beliefs (sad–happy: 12/16, $p = .077$, 95% CI [0.48, 0.93]; happy–sad: 12/16, $p = .077$, 95% CI [0.48, 0.93]); children did not infer true beliefs in the valence stable conditions (happy–happy: 9/16, $p = .804$, 95% CI [0.30, 0.80]; sad–sad: 8/15, $p = 1.000$, 95% CI [0.27, 0.79]).

Also as predicted, the effect of valence change was not significant in 4-year-olds, although there was a trend on replication, $b = -1.00$, $SE = 0.53$; $\chi^2(1, N = 61) = 3.53$, $p = .060$. Four-year-olds did not perform above chance in any condition (sad–happy: 10/16, $p = .455$, 95% CI [0.35, 0.85]; happy–sad: 11/16, $p = .210$, 95% CI [0.41, 0.89]; happy–happy: 9/14, $p = .424$, [0.35, 0.87]; sad–sad: 8/15, $p = 1.000$, 95% CI [0.27, 0.79]; see Figure 2).

Aggregating the data from the initial experiment and replication suggests that children's chance performance on the belief inferences is unlikely to be due to the experiments being underpowered: Overall, there was a main effect of the presence or absence of a valence change in 5-year-olds, $b = -1.99$, $SE = 0.41$; $\chi^2(1, N = 126) = 23.2$, $p < .001$, but not in 4-year-olds, $b = -0.51$, $SE = .36$; $\chi^2(1, N = 124) = 1.98$, $p = .160$. Five-year-olds in the valence change conditions inferred false beliefs (sad–happy: 27/32, $p < .001$, 95% CI [0.67, 0.95]; happy–sad: 25/32, $p = .002$, 95% CI [0.60, 0.91]), but chose at chance in the valence stable conditions (happy–happy: 20/31, $p = .150$, 95% CI [0.45, 0.81]; sad–sad: 19/31, $p = .281$, 95% CI [0.42, 0.78]); 4-year-olds chose at chance in all conditions (sad–happy: 18/32, $p = .600$, 95% CI [0.38, 0.74]; happy–sad: 21/32, $p = .110$, 95% CI [0.47, 0.81]; happy–happy: 17/29, $p = .458$, 95% CI [0.39, 0.77]; sad–sad: 14/31, $p = .720$, 95% CI [0.27, 0.64]).

Aggregating the data also allows us to see whether 4-year-olds' failures to infer false beliefs in the valence change conditions might be due to task switching demands. However, those 4-year-olds who saw a valence change story on the first trial were no more likely to succeed on the first trial than overall (10/17, $p = .629$, 95% CI [0.33, 0.82]).

For comparison, adults recovered agents' desires near ceiling (happy–happy: 61/61, $p < .001$, 95% CI [0.94, 1.00]; sad–happy: 60/61, $p < .001$, 95% CI [0.91, 1.00]; sad–sad: 61/61, $p < .001$, 95% CI [0.94, 1.00]; happy–sad: 61/61, $p < .001$, 95% CI [0.94, 1.00]). Comparing the adults with the 5-year-olds, there was no main effect of age, $b = 1.52$, $SE = 0.39$; $\chi^2(1, N = 370) = 2.9$, $p = .090$, but there was a main effect of the presence or absence of a valence change, $b = 2.73$, $SE = 1.51$; $\chi^2(1, N = 370) = 133.5$, $p < .001$, and an interaction, $b = -2.36$, $SE = 0.59$; $\chi^2(1, N = 370) = 15.8$, $p < .001$. Post hoc analyses found that the presence or absence of a valence change affected adults' belief inferences, $b = -4.36$, $SE = .43$; $\chi^2(1, N = 244) = 105.00$, $p < .001$. In the valence change conditions, adults and 5-year-olds both inferred false beliefs—111/122 adults versus 52/64 five-year-olds; $\chi^2(1, N = 186) = 2.83$, $p = .093$, 95% CI [−0.02, 0.22], two-sample test for equality of proportions with continuity correction; adults, sad–happy: 59/61, $p < .001$, 95% CI [0.89, 1.00]; happy–sad: 52/61, $p < .001$, 95% CI [0.74, 0.93]. However, unlike children, adults inferred true beliefs in the valence stable conditions—108/122 adults versus 39/62 five-year-olds; $\chi^2(1, N = 184) = 15.20$, $p < .001$, 95% CI [0.11, 0.40]; adults, happy–happy: 60/61, $p < .001$, 95% CI [0.91, 1.00]; sad–sad: 48/61, $p < .001$, 95% CI [0.66, 0.88].

Experiment 3

There are two ways in which Experiments 1 and 2 may have underestimated the children. First, although we said that the agent (e.g., Sally) might want an apple or might want a banana, we failed to specify that the two options were mutually exclusive, thus some of the children may have decided that she liked or disliked both. Second, because the questions were always asked in a fixed order—desire first and belief second—some of the children, and especially the youngest ones, may have had difficulty answering the second question. If so, 4-year-olds' apparent failure to infer beliefs may have been due to an overall degradation in their attention rather than a specific difficulty with belief questions. In Experiment 3, we address both these concerns by clarifying that the options are mutually exclusive and by always asking the belief question first. (We did this rather than counterbalance order both because the belief question was of primary interest given that other studies have shown that children at this age can infer agents' desires from emotional expressions and because the belief question is presumably harder given children's near ceiling performance on the desire question.) Finally, the previous results enabled us to estimate the effect size, so we ran a power analysis to ensure that we had sufficient power to detect above-chance performance (if present). The power analysis indicated that for a power of 0.80, a two-tailed alpha less than 0.05, and an effect size of Cohen's $h = 0.52$ (a conservative effect size calculated based on the replication data in the valence change conditions), we should test 29 four-year-olds and 29 five-year-olds. We preregistered this experiment on Open Science Framework (https://osf.io/nduwg/?view_only=5a9590c9183e4a01b1a4f3bcb45d6639).

Method

Participants

Fifty-eight children ($M = 5.0$ years, range: 4.0–5.9; 50% girls) were recruited from the museum between March and April 2016. To ensure a balanced distribution across ages, children were recruited in age bins consisting of 29 four-year-olds ($M = 4.5$ years, range: 4.0–4.9; 55% girls) and 29 five-year-olds ($M = 5.5$ years, range: 5.0–5.9; 45% girls).

Materials and Procedure

The materials and procedure for the children were identical to the initial experiment with two exceptions. First, when the experimenter introduced the agent's possible desires, she specified that the two candidate desires were mutually exclusive. Using the apple–banana story as an example, when the experimenter placed Picture 2 on the table, she said,

> This is Sally. She *either* likes apples *or* she likes bananas but she *doesn't* like both. Today she wants something to eat. So she might want an apple, or she might want a banana but we don't know *which one* she wants.

Second, throughout we asked the belief questions first and the desire questions second.

Coding

Coding was identical to the initial experiment. Seventy-five percent of responses were recoded by an independent coder blind to hypotheses and conditions; there was 99% agreement on children's responses. Eight of 464 responses could not be classified (e.g., "She doesn't know" and "I don't know"); as in previous experiments, these responses were coded as missing values. Additionally, two responses were dropped because of sibling interference, and two were dropped because of experimental error.

Results and Discussion

Desires

Although the desire question was asked second in this experiment, both 4- and 5-year-olds had no difficulty recovering the agents' desires. There was a main effect of the valence of the final expression, $b = -4.90$, $SE = 3.59$; $\chi^2(1, N = 227) = 89.91$, $p < .001$. The main effect of age was not significant, $b = -0.61$, $SE = 0.45$; $\chi^2(1, N = 227) = 0.00$, $p = .947$, and although the interaction between age and the valence of the final expression was significant, $b = 1.90$, $SE = 0.76$; $\chi^2(1, N = 227) = 6.26$, $p = .012$, both 5 and 4-year-olds performed near ceiling in each of the four conditions (5-year-olds—happy–happy: 27/29, $p < .001$, 95% CI [0.77, 0.99]; sad–happy: 29/29, $p < .001$, 95% CI [0.88, 1.00]; sad–sad: 23/28, $p < .001$, 95% CI [0.63, 0.94]; happy–sad: 24/27, $p < .001$, 95% CI [0.71, 0.98]; 4-year-olds—happy–happy: 28/29, $p < .001$, 95% CI [0.82, 1.00]; sad–happy: 23/28, $p < .001$, 95% CI [0.63, 0.94]; sad–sad: 24/29, $p < .001$, 95% CI [0.64, 0.94]; happy–sad: 23/28, $p < .001$, 95% CI [0.63, 0.94]).

Beliefs

Overall, we replicated children's performance. Specifically, there was a main effect of age, $b = -0.23$, $SE = 0.32$; $\chi^2(1, N = 225) = 14.06$, $p < .001$, and the presence or absence of a valence change, $b = 5.22$, $SE = 2.43$; $\chi^2(1, N = 225) = 21.80$, $p < .001$, on children's belief inferences; the interaction was also significant, $b = -1.34$, $SE = 0.49$; $\chi^2(1, N = 225) = 7.43$, $p = .006$.

Further analyses showed that 5-year-olds' inferences were affected by the presence or absence of a valence change, $b = -1.90$, $SE = 0.42$; $\chi^2(1, N = 114) = 20.50$, $p < .001$. In the valence change conditions, they inferred false beliefs (sad–happy: 21/28, $p = .013$, 95% CI [0.55, 0.89]; happy–sad: 21/28, $p = .013$, 95% CI [0.55, 0.89]); in the valence stable conditions, children inferred true beliefs in happy–happy condition but not in the sad–sad condition (happy–happy: 21/29, $p = .024$, 95% CI [0.53, 0.87]; sad–sad: 19/29, $p = .136$, 95% CI [0.46, 0.82]).

The effect of the presence or absence of a valence change was also significant in 4-year-olds, $b = -0.86$, $SE = 0.41$; $\chi^2(1, N = 111) = 4.39$, $p = .036$, but it was driven by children's unpredicted success in happy–happy condition. In the valence change conditions, 4-year-olds did not infer false beliefs (sad–happy: 11/28, $p = .345$, 95% CI [0.22, 0.59]; happy–sad: 14/28, $p = 1.000$, 95% CI [0.31, 0.69]); in the valence stable conditions, they inferred true beliefs in the happy–happy condition (24/27, $p < .001$, 95% CI [0.71, 0.98]) but not in the sad–sad condition (17/28, $p = .345$, 95% CI [0.41, 0.79]; see Figure 2).

In general, the results of Experiment 3 replicate the previous studies, suggesting that children's ability to recover both beliefs and desires from changes in the valence of agents' emotional reactions develops between ages 4 and 5. The success of both 4- and 5-year-olds at recovering the agent's beliefs in the condition when she was happy both in anticipating and observing the results was unexpected and inconsistent with the results of the previous studies. Note, however, that in this condition children could succeed simply by reporting the item actually observed in the container throughout. It is interesting that children made this response only in Experiment 3 and not in the previous studies; however, 4-year-olds' failure to recover beliefs in all of the remaining conditions suggest that their isolated success in this condition is unlikely to indicate a genuine ability to recover beliefs from the dynamics of emotional expressions. Similarly, 5-year-olds' chance performance in the sad–sad condition suggests that, at best, their ability to recover beliefs from stable emotional expressions is fragile.

General Discussion

These results suggest that by age 5 children can use changes in the valence of an agent's emotional reaction to recover both her beliefs and desires in contexts where both are unknown and the agent's actions are not differentially informative.

Four-year-olds used the emotional reactions to recover the agent's desires but did not use the valence change to infer the agent's beliefs. Moreover, neither age group reliably treated a stable valence as informative about the agent's beliefs. When someone looked happy or sad about both an expected or observed outcome, adults inferred that she expected the outcome. By contrast, children gave inconsistent responses across studies to happy expressions and consistently chose at chance in response to sad expressions.

To our knowledge, this is the first study looking at how the dynamics of facial expressions inform children's theory of mind. Previous research suggests that children selectively invoke agent's beliefs in response to surprised, curious, or frightened responses to unusual or mysterious events (Rieffe et al., 2005; Wellman & Banerjee, 1991). Here, however, and in contrast to studies where the emotional stakes have been relatively high (at least from a child's perspective—lost bunnies, dead turtles, growling dogs, etc.; Lagattuta et al., 1997; Pons et al., 2004; Widen & Russell, 2010), 5-year-olds inferred agents' beliefs given only happy and sad expressions and entirely ordinary events (e.g., finding fruit in a container). These results suggest that children treat changes in emotional valence as informative even when the emotions themselves are of relatively little import.

In some respects, the current task resembles the classic unexpected contents task (Hogrefe, Wimmer, & Perner, 1986; Perner, Leekam, & Wimmer, 1987), in that children have to reassess an initial belief about the contents of a container based on subsequent evidence. However, in the unexpected contents task, the initial belief is explicitly cued by the container itself (e.g., a "Smarties" container); the question is whether children continue to access this belief when it is subsequently proven false. By contrast, in the current task, the container provides no cues to its contents: Children must simultaneously infer both the content and epistemic status of the agent's initial belief. Thus, the current study is not merely an affective version of an unexpected contents task. Rather, it tests children's ability to use emotional expressions to infer the content of mental states that are otherwise underdetermined by the agent's actions and the context.

Like classic false belief tasks, however, this study arguably makes high demands on information processing abilities distinct from theory of mind (see Baillargeon, Scott, & He, 2010; Perner & Lang, 1999; Sodian, 2011; Wellman, 2014 for diverse perspectives and reviews). In light of this, 5-year-olds' successes in the valence change conditions may be particularly convincing, and 4-year-olds' failure less so. In particular, 4-year-olds were more successful in recovering the agent's desires than beliefs. This might be because children could infer the agent's desires in our task by attending only to a single emotional reaction (the agent's response to the observed outcome), whereas inferring beliefs required children to compare two emotional reactions (the agent's response to both the anticipated and observed outcome). Although we cannot rule out the possibility that the younger children's failure may be due to performance deficits, our results are consistent with a number of studies suggesting that children represent desires earlier and more robustly than beliefs (see Wellman & Liu, 2004 for meta-analysis). One possibility is that the processing demands are bound up with conceptual development insofar as the relative subtlety of emotional cues to beliefs versus desires may contribute to desires being represented earlier and more robustly than beliefs.

Additionally, our results are consistent with a large body of work suggesting that the integration of emotion understanding and belief–desire psychology undergoes substantial development between 4 and 6 (e.g., Bradmetz & Schneider, 1999; De Rosnay et al., 2004; Hadwin & Perner, 1991; Harris, 2008; Harris et al., 1989; Lagattuta & Wellman, 2001; Pons et al., 2004; Ruffman & Keenan, 1996; Wellman & Bartsch, 1988; Wellman & Liu, 2004; Widen & Russell, 2008). In particular, a number of studies suggest that children's ability to link mental representations of past events to current emotions improves over the preschool years. Thus, for instance, although 4-year-olds understand that someone's emotional reaction to an event may wane over time, not until 6 do children understand that this depends on whether the person remembers or forgets the event (Harris et al., 1985). Similarly, children becoming increasingly able to understand what kinds of cues about past events might trigger memories that could affect someone's emotions in the present (Lagattuta et al., 1997). In the current study, children did not have to link a mental representation of a past event to a current emotion, rather they had to use both current and past emotional expressions to infer past mental representations (the character's earlier beliefs). However, improvements in children's overall ability to represent the causal relations between mental representations and emotion over time might contribute to the developmental change in children's performance between 4 and 6 on this task.

Given that 5-year-olds recovered the agent's false beliefs from the changing valence, why, unlike adults, did they fail to attribute true beliefs consistently when the agent had the same reaction to expecting and observing an outcome? Some work suggests that children's ability to understand subtle aspects of emotion (e.g., discrepancies between true and expressed emotion, or mixed ambivalent emotions) undergoes protracted development (Pons et al., 2004). A failure to change expressions between an observed and expected outcome is prima facie, a very subtle cue to agent's mental states: Children's ability to draw inferences from such subtle cues might continue to develop through middle childhood. A related possibility is that the stable emotional expressions may have been less salient than the changing ones; thus, children might have attended less to the emotional reactions overall in the same valence conditions. Additionally, we note that some studies suggest that when children begin to pass explicit false belief tasks, they overattribute false beliefs when they should not (see Hedger & Fabricius, 2011 for a review). The current results might reflect an overall increase in children's willingness to infer false beliefs rather than true ones.

A final possibility is that 5-year-olds might have been more likely than adults to allow for the possibility that the agent liked, or disliked, both outcomes equally (even when, as in Experiment 3, the experimenter specified that the preferences were mutually exclusive). More than adults, children may have resisted learning the relatively arbitrary rule that the agent liked only one of the two items in each category given real-world experience that children who like (or dislike) fruit, desserts, snacks, or vegetables tend to apply that preference to the category as a whole. However, if the agent did like or dislike both items equally, her emotional response would be uninformative because it would be identical whether her expectations were fulfilled or violated. To the degree that children assume the agent had no preference between outcomes, they might reasonably treat the agent's stable valence as uninformative about her beliefs. In this respect, the task of jointly inferring an agent's beliefs and desires may be more difficult than inferring someone's beliefs when her desires are fully specified.

Finally, our study assumes that the agent in each story trusted the bunny's testimony and thus formed a belief about what was in the container and emotionally responded to this belief. This is consistent with other studies suggesting a strong default assumption that testimony should be trusted (e.g., Jaswal, Croft, Setia, & Cole, 2010). However, across trials, bunnies were unreliable agents—half the bunnies were correct about the contents and half were not—and previous research suggests that 4- and 5-year-olds track the reliability of informants (see, e.g., Corriveau & Harris, 2009; Jaswal & Neely, 2006; Koenig, Clément, & Harris, 2004). In this study, children did not have sufficient information to form reliability judgments given that the color of the bunny changed on each trial (suggesting that there was a different informant each time), and no individual agent had prior knowledge about that bunny's reliability. Nonetheless, we cannot rule out the possibility that children might have inferred that the bunny's testimony led the actor to represent (and emotionally react to) the contents the bunny identified without specifically developing an expectation that those contents were in the container. That is, children might have inferred that the agent formed a mental representation consisting of a "thought" about the contents rather than a "belief" about them and answered the belief question on this basis. If true, this changes our account of the results only modestly: Rather than suggesting that children used the changing valence between the expected and observed outcome to infer the agent's beliefs and desires, it would suggest that children used the changing valence between the reported and observed outcome to infer the agents' thoughts and desires.

The mental state inferences here were challenging insofar as the input was impoverished: The character showed only two expressions and did not engage in any goal-directed actions. However, the hypothesis space here was restricted to two alternatives, children had continuous access to the agent's emotional reaction to both the expected and actual outcomes, and the emotional reactions were highlighted. In the real world, emotional reactions are transient and typically go unremarked. Future research might investigate children's ability to recover agent's beliefs and desires in contexts where the emotional reactions unfold in time, and where both the hypotheses and emotional reactions are more complex than those used here. Future research might also look at older children to see when children's performance converges with adults'.

The current results, however, suggest that by age 5, children's intuitive theory of mind begins to support mental state inferences from others' emotional reactions. Extending previous work, children not only understand that thinking affects feeling (e.g., Harris et al., 1985; Lagattuta & Wellman, 2001;

Lagattuta et al., 1997; Taylor & Harris, 1983), they can use others' feelings to infer otherwise unknown thoughts. When children see someone's face change from sadness to happiness, or from happiness to sadness, they gain insight not only into how the person feels but also what she wants and believes about the world.

References

Baillargeon, R., Scott, R. M., & He, Z. (2010). False-belief understanding in infants. *Trends in Cognitive Sciences*, *14*, 110–118. doi:10.1016/j.tics.2009.12.006

Bartsch, K., & Estes, D. (1996). Individual differences in children's developing theory of mind and implications for metacognition. *Learning and Individual Differences*, *8*, 281–304. doi:10.1016/S1041-6080(96)90020-5

Bradmetz, J., & Schneider, R. (1999). Is Little Red Riding Hood afraid of her grandmother? Cognitive vs. emotional response to a false belief. *British Journal of Developmental Psychology*, *17*, 501–514. doi:10.1348/026151099165438

Corriveau, K., & Harris, P. L. (2009). Choosing your informant: Weighing familiarity and recent accuracy. *Developmental Science*, *12*, 426–437. doi:10.1111/j.1467-7687.2008.00792.x

De Rosnay, M., Fink, E., Begeer, S., Slaughter, V., & Peterson, C. (2014). Talking theory of mind talk: Young school-aged children's everyday conversation and understanding of mind and emotion. *Journal of Child Language*, *41*, 1179–1193. doi:10.1017/S0305000913000433

De Rosnay, M., Pons, F., Harris, P. L., & Morrell, J. (2004). A lag between understanding false belief and emotion attribution in young children: Relationships with linguistic ability and mothers' mental-state language. *British Journal of Developmental Psychology*, *22*, 197–218. doi:10.1348/026151004323044573

Denham, S. A., Zoller, D., & Couchoud, E. A. (1994). Socialization of preschoolers' emotion understanding. *Developmental Psychology*, *30*, 928–936. doi:10.1037/0012-1649.30.6.928

Ekman, P., & Friesen, W. V. (1976). Measuring facial movement. *Environmental Psychology and Nonverbal Behavior*, *1*, 56–75. doi:10.1007/BF01115465

Ekman, P., & Oster, H. (1979). Facial expressions of emotion. *Annual Review of Psychology*, *30*, 527–554. doi:10.1146/annurev.ps.30.020179.002523

Fabes, R. A., Eisenberg, N., McCormick, S. E., & Wilson, M. S. (1988). Preschoolers' attributions of the situational determinants of others' naturally occurring emotions. *Developmental Psychology*, *24*, 376–385. doi:10.1037/0012-1649.24.3.376

Field, T. M., Woodson, R., Greenberg, R., & Cohen, D. (1982). Discrimination and imitation of facial expression by neonates. *Science*, *218*, 179–181. doi:10.1126/science.7123230

Flom, R., & Bahrick, L. E. (2007). The development of infant discrimination of affect in multimodal and unimodal stimulation: The role of intersensory redundancy. *Developmental Psychology*, *43*, 238–252. doi:10.1037/0012-1649.43.1.238

Gnepp, J., McKee, E., & Domanic, J. A. (1987). Children's use of situational information to infer emotion: Understanding emotionally equivocal situations. *Developmental Psychology*, *23*, 114–123. doi:10.1037/0012-1649.23.1.114

Gross, A. L., & Ballif, B. (1991). Children's understanding of emotion from facial expressions and situations: A review. *Developmental Review*, *11*, 368–398. doi:10.1016/0273-2297(91)90019-K

Hadwin, J., & Perner, J. (1991). Pleased and surprised: Children's cognitive theory of emotion. *British Journal of Developmental Psychology*, *9*, 215–234. doi:10.1111/j.2044-835X.1991.tb00872.x

Harris, P. L. (2008). Children's understanding of emotion. In M. Lewis, J. M. Haviland-Jones, & L. F. Barrett (Eds.), *Handbook of emotions* (3rd ed., pp. 320–331). New York, NY: Guilford.

Harris, P. L., Guz, G. R., Lipian, M. S., & Man-Shu, Z. (1985). Insight into the time course of emotion among Western and Chinese children. *Child Development*, *56*, 972–988. doi:10.2307/1130109

Harris, P. L., Johnson, C. N., Hutton, D., Andrews, G., & Cooke, T. (1989). Young children's theory of mind and emotion. *Cognition & Emotion*, *3*, 379–400. doi:10.1080/02699938908412713

Harris, P. L., Olthof, T., Terwogt, M. M., & Hardman, C. E. (1987). Children's knowledge of the situations that provoke emotion. *International Journal of Behavioral Development*, *10*, 319–343. doi:10.1177/016502548701000304

Harwood, M. D., & Farrar, M. J. (2006). Conflicting emotions: The connection between affective perspective taking and theory of mind. *British Journal of Developmental Psychology*, *24*, 401–418. doi:10.1348/026151005X50302

Haviland, J. M., & Lelwica, M. (1987). The induced affect response: 10-week-old infants' responses to three emotion expressions. *Developmental Psychology*, *23*, 97–104. doi:10.1037/0012-1649.23.1.97

Hedger, J. A., & Fabricius, W. V. (2011). True belief belies false belief: Recent findings of competence in infants and limitations in 5-year-olds, and implications for theory of mind development. *Review of Philosophy and Psychology*, *2*, 429–447. doi:10.1007/s13164-011-0069-9

Hoehl, S., & Striano, T. (2008). Neural processing of eye gaze and threat-related emotional facial expressions in infancy. *Child Development*, *79*, 1752–1760. doi:10.1111/j.1467-8624.2008.01223.x

Hogrefe, G. J., Wimmer, H., & Perner, J. (1986). Ignorance versus false belief: A developmental lag in attribution of epistemic states. *Child Development*, *57*, 567–582. doi:10.2307/1130337

Hornik, R., & Gunnar, M. R. (1988). A descriptive analysis of infant social referencing. *Child Development*, *59*, 626–634. doi:10.2307/1130562

Hughes, C., & Dunn, J. (1998). Understanding mind and emotion: Longitudinal associations with mental-state talk between young friends. *Developmental Psychology*, *34*, 1026–1037. doi:10.1037/0012-1649.34.5.1026

Izard, C. E. (1994). Innate and universal facial expressions: Evidence from developmental and cross-cultural research. *Psychological Bulletin*, *115*, 288–299. doi:10.1037/0033-2909.115.2.288

Jaswal, V. K., Croft, A. C., Setia, A. R., & Cole, C. A. (2010). Young children have a specific, highly robust bias to trust testimony. *Psychological Science*, *21*, 1541–1547. doi:10.1177/0956797610383438

Jaswal, V. K., & Neely, L. A. (2006). Adults don't always know best preschoolers use past reliability over age when learning new words. *Psychological Science*, *17*, 757–758. doi:10.1111/j.1467-9280.2006.01778.x

Koenig, M. A., Clément, F., & Harris, P. L. (2004). Trust in testimony: Children's use of true and false statements. *Psychological Science*, *15*, 694–698. doi:10.1111/j.0956-7976.2004.00742.x

LaBounty, J., Wellman, H. M., Olson, S., Lagattuta, K., & Liu, D. (2008). Mothers' and fathers' use of internal state talk with their young children. *Social Development*, *17*, 757–775. doi:10.1111/j.1467-9507.2007.00450.x

Lagattuta, K. H., & Wellman, H. M. (2001). Thinking about the past: Early knowledge about links between prior experience, thinking, and emotion. *Child Development*, *72*, 82–102. doi:10.1111/1467-8624.00267

Lagattuta, K. H., Wellman, H. M., & Flavell, J. H. (1997). Preschoolers' understanding of the link between thinking and feeling: Cognitive cuing and emotional change. *Child Development*, *68*, 1081–1104. doi:10.1111/j.1467-8624.1997.tb01986.x

Moll, H., Kane, S., & McGowan, L. (2016). Three-year-olds express suspense when an agent approaches a scene with a false belief. *Developmental Science*, *19*, 208–220. doi:10.1111/desc.12310

Montague, D. P., & Walker-Andrews, A. S. (2001). Peekaboo: A new look at infants' perception of emotion expressions. *Developmental Psychology*, *37*, 826–838. doi:10.1037/0012-1649.37.6.826

Moses, L. J., Baldwin, D. A., Rosicky, J. G., & Tidball, G. (2001). Evidence for referential understanding in the emotions domain at twelve and eighteen months. *Child Development*, *72*, 718–735. doi:10.1111/1467-8624.00311

Mumme, D. L., & Fernald, A. (2003). The infant as onlooker: Learning from emotional reactions observed in a television scenario. *Child Development*, *74*, 221–237. doi:10.1111/1467-8624.00532

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*, aac4716-aac4716. doi:10.1126/science.aac4716

Perner, J., & Lang, B. (1999). Development of theory of mind and executive control. *Trends in Cognitive Sciences*, *3*, 337–344. doi:10.1016/S1364-6613(99)01362-5

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

Pons, F., Harris, P. L., & De Rosnay, M. (2004). Emotion comprehension between 3 and 11 years: Developmental periods and hierarchical organization. *European Journal of Developmental Psychology*, *1*, 127–152. doi:10.1080/17405620344000022

Repacholi, B. M., & Gopnik, A. (1997). Early reasoning about desires: Evidence from 14-and 18-month-olds. *Developmental Psychology*, *33*, 12–21. doi:10.1037/0012-1649.33.1.12

Rieffe, C., Terwogt, M. M., & Cowan, R. (2005). Children's understanding of mental states as causes of emotions. *Infant and Child Development*, *14*, 259–272. doi:10.1002/icd.391

Ruffman, T., & Keenan, T. R. (1996). The belief-based emotion of surprise: The case for a lag in understanding relative to false belief. *Developmental Psychology*, *32*, 40–49. doi:10.1037/0012-1649.32.1.40

Skerry, A. E., & Spelke, E. S. (2014). Preverbal infants identify emotional reactions that are incongruent with goal outcomes. *Cognition*, *130*, 204–216. doi:10.1016/j.cognition.2013.11.002

Soderstrom, M., Reimchen, M., Sauter, D., & Morgan, J. L. (2017). Do infants discriminate non-linguistic vocal expressions of positive emotions? *Cognition and Emotion*, *31*, 298–311. doi:10.1080/02699931.2015.1108904

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

Soken, N. H., & Pick, A. D. (1999). Infants' perception of dynamic affective expressions: Do infants distinguish specific expressions? *Child Development*, *70*, 1275–1282. doi:10.1111/1467-8624.00093

Sorce, J. F., Emde, R. N., Campos, J. J., & Klinnert, M. D. (1985). Maternal emotional signaling: Its effect on the visual cliff behavior of 1-year-olds. *Developmental Psychology*, *21*, 195–200. doi:10.1037/0012-1649.21.1.195

Taylor, D. A., & Harris, P. L. (1983). Knowledge of the link between emotion and memory among normal and maladjusted boys. *Developmental Psychology*, *19*, 832–838. doi:10.1037/0012-1649.19.6.832

Thompson, R. A., & Lagattuta, K. H. (2006). Feeling and understanding: Early emotional development. In K. McCartney & D. Phillips (Eds.), *The Blackwell handbook of early childhood development* (pp. 317–337). Oxford, UK: Blackwell. doi:10.1002/9780470757703.ch16

Walden, T. A., & Ogan, T. A. (1988). The development of social referencing. *Child Development*, *59*, 1230–1240. doi:10.2307/1130486

Walker-Andrews, A. S. (1997). Infants' perception of expressive behaviors: Differentiation of multimodal information. *Psychological Bulletin*, *121*, 437–456. doi:10.1037/0033-2909.121.3.437

Wellman, H. M. (2014). *Making minds: How theory of mind develops*. New York, NY: Oxford University Press. doi:10.1093/acprof:oso/9780199334919.001.0001

Wellman, H. M., & Banerjee, M. (1991). Mind and emotion: Children's understanding of the emotional consequences of beliefs and desires. *British Journal of Developmental Psychology*, *9*, 191–214. doi:10.1111/j.2044-835X.1991.tb00871.x
Wellman, H. M., & Bartsch, K. (1988). Young children's reasoning about beliefs. *Cognition*, *30*, 239–277. doi:10.1016/0010-0277(88)90021-2
Wellman, H. M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Development*, *75*, 523–541. doi:10.1111/j.1467-8624.2004.00691.x
Wellman, H. M., Phillips, A. T., & Rodriguez, T. (2000). Young children's understanding of perception, desire, and emotion. *Child Development*, *71*, 895–912. doi:10.1111/1467-8624.00198
Wellman, H. M., & Woolley, J. D. (1990). From simple desires to ordinary beliefs: The early development of everyday psychology. *Cognition*, *35*, 245–275. doi:10.1016/0010-0277(90)90024-E
Widen, S. C. (2013). Children's interpretation of facial expressions: The long path from valence-based to specific discrete categories. *Emotion Review*, *5*, 72–77. doi:10.1177/1754073912451492
Widen, S. C., & Russell, J. A. (2008). Young children's understanding of other's emotions. In M. Lewis & J. M. Haviland-Jones (Eds.), *Handbook of emotions* (3rd ed., pp. 348–363). New York, NY: Guilford.
Widen, S. C., & Russell, J. A. (2010). Differentiation in preschooler's categories of emotion. *Emotion*, *10*, 651. doi:10.1037/a0019005
Yuill, N. (1984). Young children's coordination of motive and outcome in judgements of satisfaction and morality. *British Journal of Developmental Psychology*, *2*, 73–81. doi:10.1111/j.2044-835X.1984.tb00536.x