Developmental Psychology

Integrating Expectations and Outcomes: Preschoolers' Developing Ability to Reason About Others' Emotions

Mika Asaba, Desmond C. Ong, and Hyowon Gweon

Online First Publication, May 16, 2019. http://dx.doi.org/10.1037/dev0000749

CITATION

Asaba, M., Ong, D. C., & Gweon, H. (2019, May 16). Integrating Expectations and Outcomes: Preschoolers' Developing Ability to Reason About Others' Emotions. *Developmental Psychology*. Advance online publication. http://dx.doi.org/10.1037/dev0000749

American Psychological Association

Developmental Psychology

© 2019 American Psychological Association
0012-1649/19/$12.00 http://dx.doi.org/10.1037/dev0000749

Integrating Expectations and Outcomes: Preschoolers' Developing Ability to Reason About Others' Emotions

Mika Asaba
Stanford University

Desmond C. Ong
Stanford University and A*STAR Artificial Intelligence Initiative, Agency for Science, Technology and Research, Singapore

Hyowon Gweon
Stanford University

People's emotional experiences depend not only on what actually happened, but also on what they thought would happen. However, these expectations about future outcomes are not always communicated explicitly. Thus, the ability to infer others' expectations in context and understand how these expectations influence others' emotions is an important aspect of our social intelligence. Prior work suggests that an abstract understanding of how expectations modulate emotional responses may not emerge until 7 to 8 years of age. Using a novel paradigm that capitalizes on intuitive physics to generate contextually plausible expectations, we present evidence for expectation-based emotion inference in preschool-aged children. Given two bowlers who experienced identical final outcomes (hitting 3 of 6 pins), we varied the trajectory of their balls such that one would initially expect to hit all pins (high-expectation), while the other would expect to hit none (low-expectation). In Experiment 1, both 4- and 5-year-olds appropriately adjusted characters' happiness ratings upward (low-expectation) or downward (high-expectation) relative to their initial emotions; however, only 5-year-olds made adjustments robust enough to manifest as higher final ratings for the low-expectation than the high-expectation character. In Experiments 2–3, we replicate these results and show that 5-year-olds reliably differentiate the characters' emotions even when their expectations must be inferred from context. An internal meta-analysis revealed a robust and consistent effect across the three experiments. Together, these findings provide the earliest evidence for expectation-based emotion reasoning and suggest that the ability to spontaneously generate and consider others' expectations continues to develop during preschool years.

Keywords: emotion reasoning, affective cognition, theory of mind, counterfactual reasoning, intuitive physics

Supplemental materials: http://dx.doi.org/10.1037/dev0000749.supp

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

We often find ourselves thinking about how others feel. Knowing others' emotional experiences not only allows us to sympathize, console, or comfort, but also proactively promote or prevent certain feelings in those we care about. Yet, people's emotional experiences are complex; although positive and negative events generally lead to positive and negative emotions, respectively, people's subjective responses to event outcomes can also vary depending on what they expected to happen (Mellers, Schwartz, Ho, & Ritov, 1997). Thus, to accurately infer how others feel, we must go beyond associations between the valence of event outcomes and affective states and consider how people's expectations influence their feelings.

However, people do not always broadcast what they expect, wish, or hope to happen; successful social interactions often require the ability to infer how others might feel even when explicit cues to their inner feelings are unavailable, insufficient, or misleading. One remarkable aspect of human social intelligence is that as adults, we readily make rich, flexible attributions of emotional states even from limited data. Imagine two friends at a bowling alley. Sally's ball starts off heading toward the gutter but curves back to the center to knock down three pins. Annie's ball starts rolling straight toward the center, but near the end it curves to the left to knock down three pins as well. Who feels better, Sally or Annie? In the absence of any explicit information about their expectations, one might still guess that Sally feels happier than Annie. This is because one

---

Mika Asaba, Department of Psychology, Stanford University; Desmond C. Ong, Department of Psychology, Stanford University, and A*STAR Artificial Intelligence Initiative, Agency for Science, Technology and Research, Singapore; Hyowon Gweon, Department of Psychology, Stanford University.

Mika Asaba and Desmond C. Ong contributed equally to the work.

Correspondence concerning this article should be addressed to Mika Asaba or Hyowon Gweon, Department of Psychology, Stanford University, 450 Serra Mall, Building 420, Stanford, CA 94305. E-mail: masaba@stanford.edu or hyo@stanford.edu

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

can infer the characters' expectations in context, spontaneously filling in the missing information; given the physical trajectory of the balls, Sally presumably had a much lower expectation than Annie about how many pins she would hit. This simple yet intuitive example illustrates how our understanding of the physical and the social world supports powerful inferences about others' emotional states even when event outcomes alone are insufficient and explicit information about expectations are unavailable.

A long tradition of developmental research suggests that early social cognition relies on intuitive theories of others' minds; these theories involve an abstract, causal understanding of how agents' actions and their mental states are influenced by their external environment (Carey, 1985; Flavell, 1999; Gergely & Csibra, 2003; Gopnik & Wellman, 1992; Jara-Ettinger, Gweon, Schulz, & Tenenbaum, 2016; Wellman & Woolley, 1990). Although much of the developmental literature on Theory of Mind has focused on children's understanding of goals, desires, and beliefs, some prior work has also explored how these inferences might extend to others' emotional states (e.g., Bradmetz & Schneider, 1999; De Rosnay, Pons, Harris, & Morrell, 2004; Harris, 1989; Harris, Johnson, Hutton, Andrews, & Cooke, 1989; Lagattuta, Wellman, & Flavell, 1997; Wellman & Banerjee, 1991; Wellman & Liu, 2004). Recent computational work provides a formal framework to express these ideas in more precise, quantitative terms; people can predict or explain others' emotional responses based on a generative causal model that incorporates how external events and others' internal mental states (e.g., goals, beliefs) give rise to various affective states (Ong, Zaki, & Goodman, 2015, 2018; Saxe & Houlihan, 2017; Wu, Baker, Tenenbaum, & Schulz, 2018). By characterizing emotion reasoning as an inference grounded in intuitive theories, this framework provides a principled account of how basic cognitive capacities can support a rich, flexible understanding of others' emotional states even from limited data.

The richness and the flexibility of our ability to reason about how others feel might be a hard-won developmental feat. For instance, to understand that Sally feels better than Annie, children must understand how the same outcome can elicit different emotions depending on expectations and even generate those expectations in context if they are not explicitly communicated. Although the ability to reason about others' mental states develops rapidly during preschool years (Baker, Leslie, Gallistel, & Hood, 2016; Wellman, Cross, & Watson, 2001; Wellman & Liu, 2004), how children develop an abstract, adult-like understanding of the relationship between expectations and emotions still remains an open question. The current work aims to identify early developing competence as well as developmental changes in expectation-based inferences about others' emotions.

In what follows, we first provide a brief review of prior work on young children's ability to understand others' emotions, starting from simple predictions based on event outcomes to more sophisticated inferences that require an integration of expectations and outcomes. We then motivate our main hypothesis that the ability to consider expectations in reasoning about others' emotions may already be present by late preschool years and present a series of experiments designed to test this hypothesis with 4- and 5-year-old children.

Early Development of Inferences About Emotions

Recent developmental research has revealed a surprisingly early-emerging understanding of the link between the valence of event outcomes and others' emotional reactions. Even 10-month-old infants expect an agent to express positive emotions given successful goal achievement (Skerry & Spelke, 2014), and 12-month-olds expect agents to show negative facial expressions following negative events (e.g., breaking a toy, Reschke, Walle, Flom, & Guenther, 2017). By the second year of life, children also understand that different positive experiences (e.g., observing a cool toy versus a yummy dessert) can elicit distinct positive emotional vocalizations (Wu, Muentener, & Schulz, 2017). Thus, as early as in infancy, children expect others' emotions to reflect properties of event outcomes.

During preschool years, children begin to show more abstract appreciation of the relationship between people's beliefs and emotions. For instance, 4- and 5-year-olds readily understand that someone can feel sad just from remembering past negative events (e.g., Lagattuta et al., 1997; see Lagattuta, 2014 for a review), and 3- to 5-year-olds understand that someone who expects desirable treats inside a gift box would feel happy (Harris et al., 1989). Thus, a basic appreciation of the relationship between beliefs and emotional states is present by preschool years. However, children have more difficulty when emotion inference relies on others' subjective beliefs that are inconsistent with reality (i.e., false beliefs, Bradmetz & Schneider, 1999; De Rosnay et al., 2004; Harris et al., 1989; Ronfard & Harris, 2014). Even though children, by early school years (age 6–7), have no trouble understanding that a character has a false belief, they still claim that the character would feel emotions that are consistent with reality rather than the character's false belief; for instance, even when children accurately understand that "Maxi" falsely believes that he has a full bar of chocolate (but in reality it was all gone), they would say that he feels unhappy, appealing to reality (Bradmetz & Schneider, 1999). Although these tasks require a relatively simple understanding that people's feelings should reflect their current beliefs, inferring how others feel may be more challenging when the beliefs are inconsistent with reality.

Other studies have gone beyond asking children to reason about current beliefs and current emotional states, asking instead how a character would feel upon learning that her expectation was based on a false belief (e.g., a character believes that a gift box contains candies but later learns that it actually contains stones; Harris et al., 1989; Wellman & Liu, 2004). Importantly, although children give appropriate answers by age 6, success on these tasks does not necessarily demonstrate the ability to genuinely integrate expectations and outcomes; because final outcomes are clearly desirable or undesirable, children could have responded accurately based on the valence of the outcome alone. A stronger test of the ability to integrate expectations and outcomes would involve a comparison of two characters who experience identical outcomes given different expectations, as in the bowling scenario above.

One recent study provides a direct test of this ability by looking at how children rate the happiness of different characters based on their prior expectations and actual outcomes (Lara, Lagattuta, & Kramer, 2017). Consistent with prior literature, their results suggest that the ability to integrate prior expectations with unexpected outcomes develops relatively late in childhood. For instance, 8- to

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

10-year-old (but not younger) children appropriately rated that a character with high expectations (e.g., someone who expects to win a big teddy bear in a raffle) would feel worse than a character with low expectations (e.g., someone who expects to not win a big teddy bear), following the same unexpected outcome (i.e., both characters win a small teddy bear).[1] Another recent study finds that 6-year-old children show some limited competence; given two agents who drew the same gumball from two different gumball machines, children understand that the one who drew a low-probability gumball feels more surprised than the one who drew a high-probability gumball (Doan, Friedman, & Denison, 2018). However, even though children had clear visual representations of agents' expectations (pictures of two gumball machines showing distinct distributions), 6-year-olds required explicit instruction to consider each agent's chances of getting the gumball to make successful judgments of the agent's emotions; 5-year-olds were unable to make appropriate emotion inferences even with encouragement.

Why So Difficult? Lack of Competence Versus Task Demands

Collectively, prior literature suggests that the ability to integrate beliefs and outcomes to infer emotions may emerge gradually during early school-aged years. Younger children may rely primarily on the valence of event outcomes or beliefs to infer others' emotions, and an adult-like understanding of how expectations shape future emotions may not be robust until 7 or 8 years of age. However, such late-emerging success is somewhat surprising given that children, by the end of preschool years, readily infer emotions from outcomes (Skerry & Spelke, 2014; Wu et al., 2017) and emotions from beliefs (Wellman et al., 2001; Wellman & Liu, 2004). Given that they already show the prerequisite inferential abilities, why is it so difficult to consider both expectations and outcomes to reason about others' feelings?

One possibility is that this reflects a genuine limitation in preschoolers' representational and inferential capacities. Note that studies that have found early competence for emotion inferences have used tasks where agents' beliefs are either unconfirmed or consistent with reality, such that they are expected to experience emotions that are consistent with their beliefs (e.g., Harris et al., 1989; Lagattuta et al., 1997). However, just as inferring beliefs is harder when it is inconsistent with reality or children's own knowledge (Wellman et al., 2001; Wimmer & Perner, 1983), inferring emotions may be harder when it involves agents' expectations that are incongruent with reality (Lara et al., 2017; see also Bradmetz & Schneider, 1999; De Rosnay et al., 2004; Harris et al., 1989; Ronfard & Harris, 2014). Beyond belief-based emotion reasoning (e.g., Bradmetz & Schneider, 1999; De Rosnay et al., 2004; Harris et al., 1989), children's performance on more complex social reasoning tasks, such as moral reasoning and pragmatic inference, also continues to improve throughout middle to late childhood (e.g., Cushman, Sheketoff, Wharton, & Carey, 2013; Filippova & Astington, 2008; Gweon, Dodell-Feder, Bedny, & Saxe, 2012; Lackner, Bowman, & Sabbagh, 2010; see also Harris, 1989). Prior work also finds that the ability to understand feigned or disguised emotions develops late (Gnepp & Hess, 1986), and even 7-year-olds have trouble understanding that someone who acted on a negative intent would feel "good" after a bad outcome that fulfills her intent (Yuill, 1984). Thus, it is possible that an adult-like understanding of how expectations modulate emotions does not emerge until age 6 or later, and preschool-aged children may lack the competence to consider unfulfilled expectations to infer how others feel.

Another possibility, however, is that these failures do not necessarily reflect a genuine lack of competence; to the extent that they are capable of belief-based reasoning (i.e., 4- and 5-year-olds, see Baker et al., 2016; Wellman et al., 2001), children may also have the underlying competence to infer others' emotions given beliefs or expectations, but their competence may easily fail to manifest due to various task demands. One hint comes from recent work (Doan et al., 2018) that capitalized on children's understanding of intuitive statistics (Denison, Trikutam, & Xu, 2014; Gweon, Tenenbaum, & Schulz, 2010). Rather than using scenarios where characters hold expectations that are somewhat arbitrary (e.g., a child expects to win or not win a big teddy bear at a raffle, without clear reasons for believing so; see Lara et al., 2017), this study found limited competence in 6-year-olds by using a task where expectations were reasonable given the relative proportion of two objects (e.g., a child expects to draw a red gumball from a machine that has many more red than black gumballs). These results suggest that children might find it easier to consider expectations about future outcomes when the expectations are causally plausible based on observed evidence.

Indeed, the presence of early competence is not inconsistent with the possibility that the basic competence for expectation-based emotion reasoning continues to develop during preschool years and beyond. A recent study suggests that 5-year-olds can infer others' beliefs and desires from changes in their emotional expressions before and after a final outcome (Wu & Schulz, 2018); 4-year-olds, however, readily infer desires but fail to infer beliefs, suggesting both early emerging competence and developmental change in young children's ability to infer mental states from emotions. Thus, children's capacity to infer emotion from mental states may show similar developmental change during this period.

Using Intuitive Physics as Contextual Support for Affective Reasoning

The current study aims to address these possibilities by investigating the early development of expectation-based emotion inference. Specifically, we hypothesize that given clear contextual support that grounds agents' expectations in concrete, physical events, even preschool-aged children might understand that two agents with different expectations can feel differently about the same outcome. To test this hypothesis, we designed a task similar to the bowling scenario above. Even infants expect an object in motion to continue on a continuous, linear path (Kotovsky & Baillargeon, 2000; Spelke, Breinlinger, Macomber, & Jacobson, 1992; Spelke, Katz, Purcell, Ehrlich, & Breinlinger, 1994) and understand that agents' goal-directed actions can set an inert object in motion or alter physical states of the world through causal interventions (Gopnik, Sobel, Schulz, & Glymour, 2001; Muent-

---

[1] While 6- and 7-year-olds did show this distinction following a negative outcome (not winning anything), it is possible that children rated the low-expectation character as happier because his expectation was more consistent with the actual outcome.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

ener & Carey, 2010; Newman, Keil, Kuhlmeier, & Wynn, 2010; Saxe, Tenenbaum, & Carey, 2005). Thus, it is possible that 4- and 5-year-old children can infer others' expectations about the outcomes of physical events given dynamic, visual information that supports their reasoning. Furthermore, tasks that capitalize on intuitive physics may be more effective than sampling events used in prior work (Doan et al., 2018); the outcomes of physical, causal events are arguably more deterministic than outcomes of sampling events, and children might find it easier to use them as support for further inferences.

We implemented the bowling scenario in a novel, minimally verbal paradigm and asked whether even 4- and 5-year-olds can successfully infer Sally and Annie's emotional states after seeing identical outcomes that violate their initial expectations. In addition to identifying early competence, we also explored potential developmental changes based on patterns observed in prior work (e.g., Doan et al., 2018; Wu & Schulz, 2018).

In Experiment 1, we ask whether 4- and 5-year-olds, as well as adults, provide appropriate emotion ratings to the two bowlers depending on different initial trajectories of their balls, given explicit information about their expectations. In Experiment 2, we remove explicit information about the agents' expectations and ask whether children can spontaneously consider others' expectations in context. In Experiment 3, we replicate findings from Experiment 2 and rule out the possibility that children are simply using their own expectations to reason about the agents' emotions. Finally, we conduct a mini meta-analysis across Experiments 1–3 to determine a more precise estimate of the effect size for 5-year-olds and to examine differences between 4- and 5-year-olds. All experiments, data, and analyses can be found at https://osf.io/wakbu/.

Experiment 1

In Experiment 1, we asked whether 4- and 5-year-old children can use explicit information about agents' expectations to infer the direction of change in agents' emotions after an outcome is revealed. Specifically, we asked whether children understand that a character with low expectations about the outcome (low-expectation trial) would feel better after the final outcome is revealed and that a character with high expectations about the outcome (high-expectation trial) would feel worse after the final outcome. If children can appropriately predict the change in characters' emotions, this would suggest that children can use others' prior expectations to inform their emotion inferences even when these expectations are no longer consistent with reality. Given prior work that shows marked difference in belief-based emotion reasoning between 4- and 5-year-olds, we planned to look at the two age groups separately. We also recruited adult participants to compare their responses to children's evaluations.

Method

Participants. We recruited 36 4- and 5-year-old children from a university preschool. The majority of children were from middle-class families with diverse cultural and ethnic backgrounds that were representative of the local population. For this initial exploratory study, we set the sample size based on convention, with an a priori decision to look at 4- and 5-year-olds separately (4-year-olds: $n = 18$, eight females, $M_{\text{Age}} = 4.51$, $SD = .32$, range = 4.00–4.97; 5-year-olds: $n = 18$, nine females, $M_{\text{Age}} = 5.43$, $SD = .27$, range = 5.05–5.98). An additional 10 children (7 4-year-olds; 3 5-year-olds) were excluded due to incorrect responses to the initial emotion judgment (i.e., reporting the low-expectation character to be happy or the high-expectation character to be sad; see Procedures).

We also recruited 17 adult participants through Amazon Mechanical Turk (eight females, $M_{\text{Age}} = 31.58$, $SD = 5.92$, range = 23–43). One additional participant was tested but excluded due to incorrect responses to a check question. All studies were approved by the Stanford University Internal Review Board (Project Name: "Research on Learning and Social Cognition In Infants, Children, and Adults"; Protocol Number IRB-31350).

Materials. Children played a warm-up game using a child-friendly toy bowling set from Kidoozie Toys: Six colorful foam pins, a red foam ball, and a black tarp for the bowling lane. Stimuli in the main task were animated video clips presented in Keynote on a 13-in. Apple MacBook Pro. Each video was approximately 12 seconds long. The cartoon bowling lane in the videos consisted of a wide beige rectangle flanked by two narrow gray rectangles (the "gutter"; see Figure 1); we designed the cartoon pins and balls to match the bowling set that children had played with initially. We used simple, generic cartoon characters for the practice and test trials. In all trials, the characters' backs were facing the child such that no facial expressions were shown. The same stimuli were used with adults.

To measure children's inferences about changes in emotion, we constructed a rating scale (approx. 10 × 2 inches.) with arrows at each end. The scale was divided into eight equal sections; a red sad face was placed on the third section and a green happy face was placed on the sixth section to serve as initial "anchor points" (see Figure 2).

Procedures.

Task introduction. Children were tested in a quiet room. To ensure that children were familiar with bowling, the experimenter introduced the game by explaining that the goal is to roll the ball to knock down as many pins as possible; children played with the bowling set for approximately 2–4 minutes. The experimenter then introduced the bowling game on the computer. First, children were shown an image of the bowling lane and pins without any characters present; they were asked how many pins were on the screen (correct answer: "six") and what it meant when the bowling ball entered the gray areas (correct answer: "it's out"). If a child could not count correctly or did not remember the answer, the experimenter repeated the explanation. This procedure ensured that all children understood the goal of the bowling game, understood the task, and correctly reported the number of pins on the screen.

Practice trials. Children first saw two practice trials (Gutter and Strike, fixed order). In the Gutter trial, the character rolled the ball, which curved to the right into the gutter and knocked down zero pins. In the Strike trial, the character's ball went straight and knocked down all six pins. After each trial, the experimenter first asked the child how many pins were knocked down (memory check). Then the experimenter asked whether the character is feeling happy or sad, and participants responded by placing a small marker on either the happy or the sad face on the scale.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

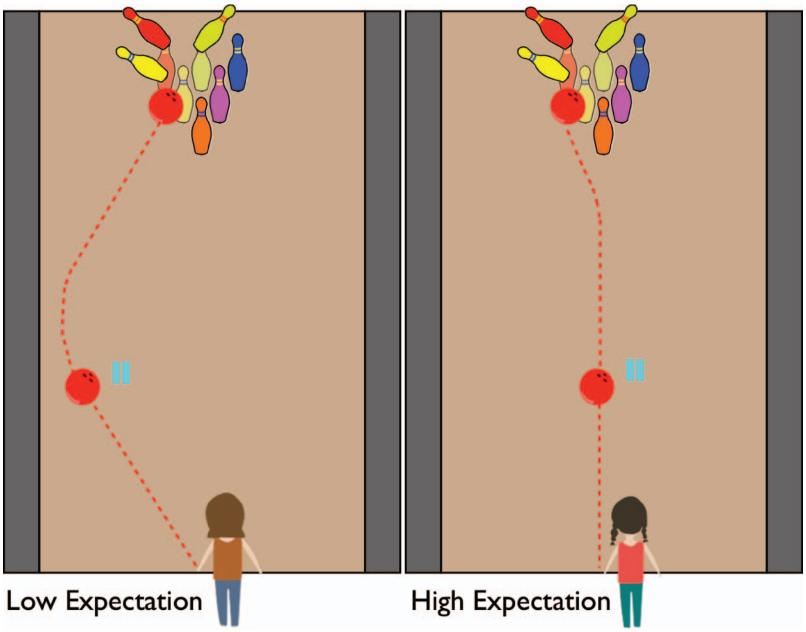

Figure 1. Sample stimuli used for Experiments 1–3. Left: Low-expectation video; right: High-expectation video. Trajectories of the bowling balls are indicated by the dashed red lines. For Experiments 1 and 3, the video was paused midway; the cyan "pause" symbol indicates the position of the bowling balls at the pause. See the online article for the color version of this figure.

Test trials. Children then saw the low-expectation and high-expectation test trials in a counterbalanced order. In the low-expectation trial, the ball initially headed left toward the gutter, and then the movie was paused partway through. The experimenter provided explicit information about the character's belief ("Sally thinks that her ball is going to go out and hit none of the pins!"), and children rated how the character feels right now ("happy or sad") by placing a marker on one of the two anchor points on the scale (initial emotion judgment). Once the movie resumed, the ball curved back to hit three pins. In the high-expectation trial, the ball initially headed straight toward the pins. The movie paused, the experimenter provided explicit information about the character's belief ("Annie thinks that her ball is going to go straight and hit all of the pins!"), and the child made an initial emotion judgment; after the pause, the ball instead curved to the left to hit three pins. Thus, although the actual outcome was identical across trial types, the initial trajectory clearly differed, and children were given explicit information about the characters' initial expectations about the outcome, which were consistent with the initial trajectories of the ball. As we were mainly interested in how children adjust their initial ratings after the outcome, children who inaccurately answered this question were excluded from analyses.

After each movie, children were first asked how many pins were knocked down (memory check; correct answer was "3" for both trial types). The experimenter then reminded the child of their earlier emotion rating, and asked them to adjust their ratings. She said: "Earlier you said she was feeling happy (sad). Do you think she is feeling better or worse now?" In addition to a verbal response, the experimenter asked the child to provide a final emotion rating for the character by saying, "Okay, can you show me? You can move her anywhere on the line." The marker's final position on the scale was used as the final rating. Finally, after both trials, children were asked to choose the character who feels happier.

Adult experiment. Adults participated in a similar procedure online, though they were not given the practice trials. Similar to the child study, the test movie was paused partway through; participants were given the character's belief about the trajectory ("Annie thinks her ball is going to go straight/out"). They were prompted to predict how many pins the character expects to knock down (free-response from 0 to 6) and then answered the initial emotion judgment question ("happy" or "sad") during the pause. After the final outcome, participants indicated whether the character feels better or worse and provided a final emotion rating for the character using a scale similar to the one given to children. Participants were then asked how many pins the character knocked down (memory check) and those who failed to answer accurately were excluded from analyses. Finally, after both trials, participants were asked to choose who is happier.

Results

We first looked at adult participants' responses. All adults correctly reported the initial emotion for each character ("sad" for the low-expectation and "happy" for the high-expectation character). Our main interest was whether they correctly predicted how their initial emotions would change after the final outcome.[2] They raised their initial emotion ratings for the low-expectation character (final – initial rating: $M = 3.29$, $SD = .92$), $t(16) = 14.77$, $p < .001$, and lowered their ratings for the high-expectation character

---

[2] Unless noted otherwise, we used paired t tests for ratings and binomial tests for binary responses.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

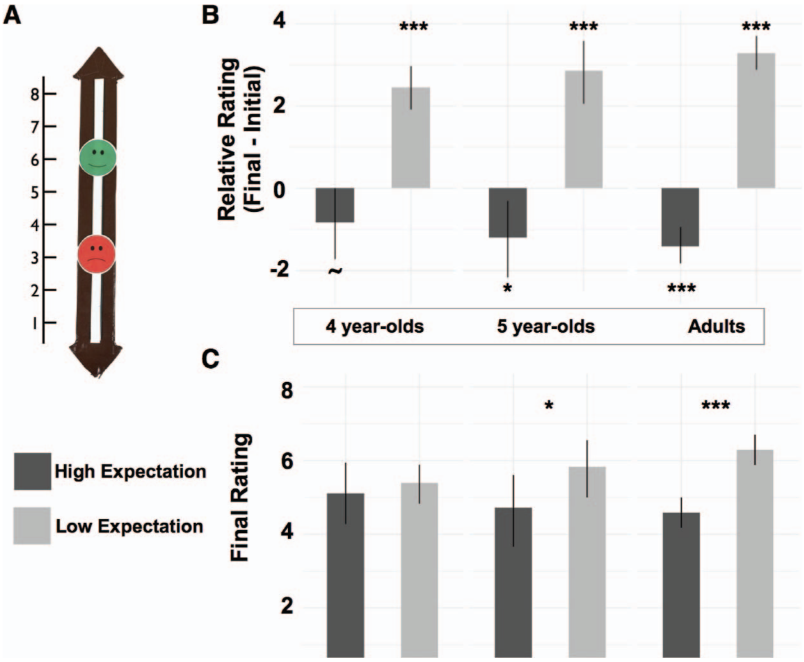

Figure 2. Results from Experiment 1. (A) Scale used for emotion ratings. (B) Participants' relative ratings (final – initial rating) for each character's emotions. (C) Participants' final ratings after the outcome. Error bars represent 95% confidence intervals. * $p = .05$. *** $p < .001$. $\sim$ $p < .1$. See the online article for the color version of this figure.

(final – initial rating: $M = 1.4$, $SD = .94$), $t(16) = -6.20$, $p < .001$. Interestingly, there was an asymmetry in how much each character's emotion rating was adjusted: Participants showed a greater magnitude in the change of ratings for the low-expectation than the high-expectation character, low-expectation versus high-expectation adults, $t(16) = 6.11$, $p < .001$.[3] For the final ratings, adults rated the low-expectation character significantly higher than the high-expectation character (low-expectation versus high-expectation: 6.30 versus 4.59, $t(16) = 5.18$, $p < .001$). Finally, 100% of participants chose the low-expectation character as feeling happier ($p < .001$).

We then looked at children's responses. During the practice trial, all children in both age groups chose the Strike character as happier than the Gutter character, suggesting that they understood the task instructions. As our main question was whether children can understand how the characters' emotion might change following the final outcome, children who incorrectly reported the low- and high-expectation characters' initial emotions were excluded from further analyses ($n = 10$; see Participants); so, participants included in analyses all had an initial rating of "3" for the low-expectation character and "6" for the high-expectation character.

Children's responses were generally consistent with those from adults. Five-year-olds raised the rating for the low-expectation character (indicating that the character felt happier; final – initial rating: $M = 2.86$, $SD = 1.66$, $t(17) = 7.31$, $p < .001$), and lowered it for the high-expectation character (final – initial rating: $M = -1.19$, $SD = 2.12$, $t(17) = -2.40$, $p = .028$). Four-year-olds showed a similar but somewhat weaker pattern; the upward change for the low-expectation character was significant (final – initial rating: $M = 2.44$, $SD = 1.21$, $t(17) = 8.56$, $p < .001$), but the downward change for the high-expectation character was marginally significant (final – initial rating: $M = -.83$, SD 1.91, $t(17) = -1.85$, $p = .082$). Mirroring the pattern in adults, the upward adjustment for the low-expectation character was larger than the downward-adjustment for the high-expectation character for both age groups (low-expectation versus high-expectation, 5-year-olds: $t(17) = 2.46$, $p = .025$; 4-year-olds: $t(17) = 2.3$, $p = .034$).[4]

For the final ratings, 5-year-olds rated the low-expectation character as happier than the high-expectation character (low-expectation versus high-expectation: 5.83 versus 4.72, $t(17) = 2.10$, $p = .051$). However, although 4-year-olds' were able to raise or lower their emotion ratings in appropriate directions, the magnitude of the changes were not robust enough to manifest as a significant difference in the final ratings (low-expectation versus high-expectation: 5.39 versus 5.11, $t(17) = .46$, $p = .654$). Additionally, neither group was able to explicitly report that the low-

---

[3] This pattern was also reflected in their binary responses; all reported that the low-expectation character feels better after seeing the outcome (100%, $p < .001$) but only 76.5% said that the high-expectation character feels worse after the outcome ($p = .049$), though this difference was not significant.

[4] Again, this pattern was reflected in children's binary responses; the vast majority of children in both age groups reported that the low-expectation character felt better after the outcome than at the pause (4-year-olds: 88.89%, $p = .001$; 5-year-olds: 94.44%, $p < 0.001$), but they were at chance for reporting the high-expectation character as feeling worse after the outcome (4-year olds: 66.67%, $p = .238$; 5-year olds: 55.56%, $p = .815$).

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

expectation character feels happier than the high-expectation character (% choosing low-expectation character: 4-year-olds: 55.56%, $p = .815$; 5-year-olds: 66.67%, $p = .238$).

Discussion

Given explicit information about characters' expectations, all age groups, including 4-year-olds, were able to change their initial emotion judgments in appropriate directions after observing the final outcomes. Given the same unexpected outcome, participants adjusted their ratings down for the high-expectation character and raised them up for the low-expectation character. This provides the earliest evidence for preschoolers' ability to consider expectations to infer others' emotions given matched outcomes.

One might wonder whether these results simply reflect a "regression to the mean"; that is, given that the high-expectation character was considered "happy" initially and the low-expectation character was considered "sad," children might have simply wanted to adjust their rating closer to the global mean. However, the asymmetry between participants' upward and downward adjustments suggests that this explanation is unlikely; in both children and adults, the magnitude of the upward change for the low-expectation was consistently greater than the downward change for the high-expectation character. Ratings based solely on the final outcomes and/or a simple regression to the mean cannot explain this pattern.

A stronger evidence for the ability to consider prior expectations comes from comparing the final ratings between the two characters. The results suggest that both adults and 5-year-olds understand that the low-expectation character feels happier than the high-expectation character. However, 4-year-olds' adjustment of emotion ratings were not robust enough to show this difference. Even though a majority of 4-year-olds adjusted the ratings in the right direction, their downward adjustment of emotions for the high-expectation character was only marginally different from zero. Thus, these findings provide only suggestive evidence that 4-year-olds consider prior expectations to adjust their emotion ratings.

Note however that in this study, all participants were given explicit information about the character's expectations and rated the characters' emotions twice, both before and after the final outcome. Even though 5-year-olds' ability to differentiate the two characters' final emotions is impressive, it remains unclear whether their success reflects a genuine ability to infer others' emotions by generating others' expectations spontaneously in context.

In sum, results from Experiment 1 provide preliminary evidence that preschool-aged children understand how prior beliefs modulate others' emotional responses. In the next two experiments, we seek to replicate the difference in the final ratings and ask whether children can discriminate the two agents in their final ratings even in the absence of explicit information about their expectations.

Experiment 2

In Experiment 2, we examined young children's and adults' ability to consider others' expectations to infer their emotions in the absence of any explicit mention of the characters' expectations. Participants viewed the same stimuli as in Experiment 1 but without any pause and provided only a final rating for each character.

Unlike Experiment 1 where information about the expected outcome was given to the participants, one important prerequisite for success in Experiment 2 was the ability to use the initial trajectory of the ball to predict the likely outcome of the event. To verify that children can predict where the ball would go given its initial trajectory, we ran a separate experiment with children ($n = 40$, age range = 4.0–5.9) where we paused the videos halfway through and simply asked children to predict where the ball was going to go. The majority (90%, $p < .001$) of 4- and 5-year-olds correctly reported the ball's likely trajectory (for detailed procedure and results, see online supplemental materials). Thus, if children fail to distinguish the two characters' emotions in Experiment 2, their failure would likely reflect their difficulty with emotion reasoning rather than a failure with physical reasoning.

Method

Participants. We aimed for a similar sample size as Experiment 1. Therefore, we recruited 40 children from a university preschool (4-year-olds: $n = 20$, eight females, $M_{\text{Age}} = 4.53$, $SD = .26$, range = 4.20–4.99; 5-year-olds: $n = 20$, 13 females, $M_{\text{Age}} = 5.32$, $SD = .17$, range = 5.06–5.61). One additional child was excluded from analysis due to failure to respond to test questions. Participants were representative of the socioeconomic and ethnic backgrounds of the local population. We also recruited 20 adults (seven female, $M_{\text{Age}} = 32.8$, $SD = 8.06$, range = 22–59) through Amazon's Mechanical Turk; no one failed the memory check.

Materials. Children played with the same child-friendly toy bowling set as in Experiment 1, and the stimuli were the same.

Procedures. The procedures were similar to Experiment 1. One key difference was that children watched the movies without any pause or prompts, and answered questions at the end of each movie. Given that it was unnecessary to rate their emotions twice, we also simplified the final question; instead of asking children to place a marker on a physical scale, children were asked whether the character is feeling happy or sad, followed by a choice between "kind of happy (sad), medium happy (sad), or really happy (sad)?" This allowed us to convert children's responses on a 6-point scale, from 1 (*really sad*) to 6 (*really happy*). After both practice trials, the experimenter asked: "Who do you think is happier?" If the child did not respond or said both, the experimenter asked the child to select one character.

Adults participated in an online experiment nearly identical to Experiment 1, but without any pause during the movies. For each character, they rated how the character feels on a 6-point Likert scale from 1 (*really sad*) to 6 (*really happy*).

Results

All participants (both children and adults) correctly reported the number of pins knocked down in every trial, and there were no significant effects of trial order or character type. Consistent with the results in Experiment 1, adults provided higher ratings for the low-expectation character than for the high-expectation character (low-expectation versus high-expectation: 4.35(.99) versus 3.70(.80), $t(19) = 3.32$, $p = .004$), and chose this character as the "happier" of the two (% choice for low-expectation character:

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

100%, $p < .001$), even without any information about the characters' initial expectations.

Next, we looked at children's responses and asked whether they made reasonable ratings for the characters in the practice trials. As in Experiment 1, all reported that the Strike character feels happier than the Gutter character. This was reflected in their ratings; both age groups provided higher ratings to the Strike character than the Gutter character, 4-year-olds (Strike versus Gutter $M(SD)$: 5.95(.22) versus 1.30(.57), $t(19) = 35.42$, $p < .001$; 5-year-olds: Strike versus Gutter: 5.95(.22) versus 1.70(.57), $t(19) = 29.76$, $p < .001$).

Our main question was whether children distinguished the high- and low-expectation characters; unlike Experiment 1, children had to infer the character's initial expectation in context based on the observed trajectory of balls, and incorporate the expectation to their emotion rating after the final outcome. The results largely replicated our findings in Experiment 1. As expected, 5-year-olds rated the low-expectation character higher than the high-expectation character (low-expectation versus high-expectation: 4.80(.95) versus 4.30(1.22), $t(19) = 2.13$, $p = .047$), while 4-year-olds did not differentiate the two characters (low-expectation versus high-expectation: 4.70(1.34) versus 4.85(1.14), $t(19) = -.55$, $p = .591$; see Figure 3). However, neither age group showed a clear preference in the binary choice question (4-year-olds: 45%, $p = .824$; 5-year-olds: 65%, $p = .263$).

Discussion

The results from Experiment 2 were consistent with our findings in Experiment 1; adults and 5-year-olds, but not 4-year-olds, gave higher emotion ratings to the low-expectation character than the high-expectation character. Remarkably, they did so even in the absence of explicit information about their expectations, suggesting that they were able to generate characters' expectations in context based on the physical trajectory of the balls. These results provide stronger support for our hypothesis that by age 5, children can consider others' expectations as well as outcomes of events to infer how they feel after an unexpected outcome.

By contrast, 4-year-olds were unable to show this distinction. This is unlikely to be due to their simple failure to predict the appropriate outcome of the event based on the initial trajectory of the ball (see online supplemental materials); it is possible that even though 4-year-olds have a nascent understanding of how beliefs and outcomes modulate others' emotions (especially upward given a low expectation), it may not be robust enough to manifest as a clear difference between the two characters' final emotions.

One might wonder whether 5-year-olds' emotion ratings reflect a genuine understanding of others' emotions. More specifically, one alternative explanation for 5-year-olds' success is that children simply relied on their own expectations about the outcome to infer the characters' emotions. Because children made the ratings after watching each movie for the first time, the outcomes were unexpected not only for the characters but also for children themselves; thus, children could have succeeded in the task without ever having to attribute expectations to the characters. In Experiment 3, we address this alternative possibility.

Experiment 3

Experiment 3 was designed with three goals. First, we wanted to address the possibility that children were relying on their own expectations about the outcome to make the emotion judgments. Here, children viewed each movie twice and provided emotion ratings after the second viewing; thus, they were fully aware of the final outcome and needed to reason about the characters' expectations to appropriately judge their emotions.

Second, we sought to better understand why 4-year-olds failed to distinguish the two characters in both Experiments 1 and 2. Our supplementary experiment rules out the possibility that they lack the basic ability to predict the appropriate outcome of the physical event; yet, a remaining possibility is that children still have difficulty attributing appropriate expectations to the characters or combining expectations and outcomes to make the final emotion rating. In Experiment 1, 4-year-olds struggled with the task even with explicit information about the characters' beliefs. Furthermore, prior work suggests that directly probing children's belief attribution before emotion inference can unexpectedly hinder their performance (Doan et al., 2018). Thus, in Experiment 3, we asked whether children can make appropriate initial emotion judgment without any information about the characters' beliefs; if 4-year-olds successfully make this judgment but still fail to distinguish the two characters' final emotions, this would suggest that their difficulty comes from the inability to consider both expectations and outcomes, rather than their inability to attribute expectations.

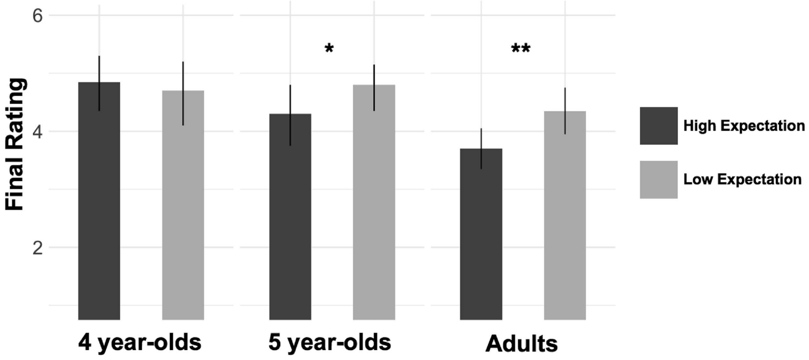

Figure 3. Results from Experiment 2. Participants' ratings of the characters' emotions after the test trials. Error bars indicate 95% confidence intervals. * $p < .05$. ** $p < .01$.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Finally, this design allowed us to replicate the results again with a higher-powered sample and a more objective dependent measure; rather than using the verbal rating method in Experiment 2, we used a physical 6-point scale to get a single final rating after the final outcome.

Method

Participants. For Experiment 3, we determined the sample size before data collection using results from Experiment 2 (effect size Cohen's $d = .475$, $n = 30$ required for 70% power). We recruited 32 4-year-olds (20 females, $M_{\text{Age}} = 4.52$, $SD = .30$, range = 4.01–4.98) and 32 5-year-olds (15 female, $M_{\text{Age}} = 5.33$, $SD = .25$, range = 5.01–5.97) from a university preschool. Five additional children were excluded from analyses due to failure to respond to test questions ($n = 1$), incorrect responses to check questions ($n = 3$), or technical error ($n = 1$). We did not recruit any adults for this study.

Materials. During the Task Introduction phase, we used a toy bowling set (same as Experiment 1) and a 30-s animated clip ("Piper" from Pixar). The videos with pauses in the main task were identical to Experiment 1 (as shown in Figure 1). For emotion ratings, we constructed a 6-point scale with smiley faces, each face representing a point on a scale from 1 (*really sad*) to 6 (*really happy*).

Procedure. The procedure was similar to Experiment 2 but with a few key changes. First, to ensure that even the youngest children understood the concept of pausing a movie, children first watched a 30-s animated clip as the experimenter paused it twice and explained what it means to "pause." Children then watched the entire clip again with no pauses. Children also received a brief training with the emotion rating scale. The experimenter said, "We are going to watch some of my friends play the game! We are going to use this (pointing to the scale) for guessing how the character feels." She then explained what each face on the scale meant (e.g., "If you think the character feels 'really sad,' you can put the marker down here"). Then the experimenter asked the child to point out where to put the marker if the character felt "really happy," "medium happy," "kind of happy," "kind of sad," "medium sad," and "really sad."

During the test trials, children saw each video clip twice. During the first viewing, the video was "paused" midway through; to make the expected outcome as clear as possible, the pause occurred just before the ball changed directions. Figure 1 shows the frames from the two test trial videos at the time of the pause (the cyan icon and red dashed line did not appear in the actual videos). After pausing the movie, the experimenter asked, "How does Annie (Sally) feel right now? Happy or sad?" (initial emotion question). Children responded verbally with "happy" or "sad," and then the video was resumed to play until the end of the clip. Critically, children then watched the same video for a second time without the pause, and the experimenter asked the child to indicate how the character was feeling using the rating scale (final emotion rating). After observing both test trials, children were asked which agent felt happier.

Results

As in Experiments 1 and 2, almost all children successfully reported that the Strike character felt "happy" (4-year-olds: 100%, $p < .001$; 5-year-olds: 100%, $p < .001$) and that the Gutter character felt "sad" (4-year-olds: 90.63%, $p < .001$; 5-year-olds: 96.88%, $p < .001$).

First, we looked at children's responses to the initial emotion question. Both 4- and 5-year-olds understood clearly that the high-expectation character felt "happy" at the pause (4-year-olds: % responding "happy": 90.63%, $p < .001$; 5-year-olds: 90.63%, $p < .001$) and the low-expectation character felt "sad" at the pause (4-year-olds: % responding with "sad": 68.75%, $p = .050$; 5-year-olds: 93.75%, $p < .001$). Post hoc comparison showed that 4-year-olds were less likely than 5-year-olds to report that the low-expectation character felt "sad" at the pause (68.75% versus 93.75%, $p = .022$, Fisher's exact test).

We then looked at the final emotion ratings. Replicating our findings in Experiment 1 and 2, 5-year-olds provided higher ratings to the low-expectation character (low-expectation versus high-expectation: 4.63(.94) versus 4.00(1.19), $t(31) = 2.797$, $p = .009$), while 4-year-olds did not distinguish the low-expectation and high-expectation characters (low-expectation versus high-expectation: 4.50(1.27) versus 4.50(1.24), $t(31) = 0$, $p = 1$). See Figure 4.

Even among the 4-year-olds who accurately responded to both initial emotion questions ($n = 19$), we did not find evidence for a difference in their final ratings, low-expectation versus high-expectation (4.63(1.27) versus 4.53(1.24), $t(18) = .309$; $p = .761$). Consistent with earlier experiments, however, neither age group explicitly chose the low-expectation character as the happier of the two (% choosing low-expectation character: 4-year-olds: 53.13%, $p = .860$; 5-year-olds: 62.5%, $p = .215$).

Discussion

These results largely mirrored our earlier findings in Experiments 1 and 2. Importantly, 5-year-olds' success in this experiment rules out the alternative explanation that children were simply relying on their own expectations about the outcomes; because children were fully aware of the final outcomes by the beginning of the second viewing, the results suggest that 5-year-olds made their emotion ratings based on what the character (or someone without any knowledge about the actual outcome) would have expected, rather than what they themselves expected to happen.

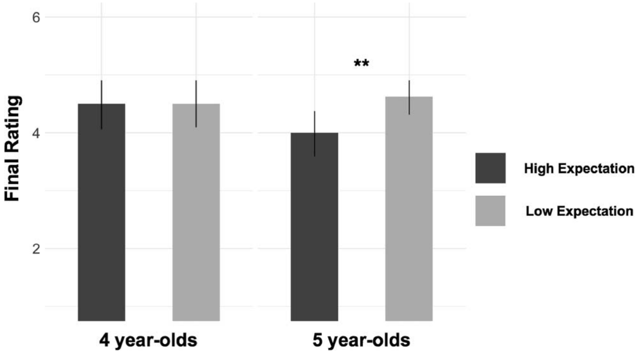

Figure 4. Results from Experiment 3. Participants' ratings of the characters' emotions after the final outcome. Error bars represent 95% confidence intervals. ** $p < .01$.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Four-year-olds showed reasonable success on the initial emotion question, but still had difficulty attributing different final emotions to the two characters. The 4-year-olds' difficulty is especially compelling given that even those who successfully answered the initial emotion questions showed no differences in their final emotion ratings. This suggests that the ability to understand the valence of expected outcomes and mapping them onto corresponding affective states (i.e., attributing expectations and linking them to emotions) is insufficient for successfully inferring how these emotions might change after an unexpected outcome.

Mini Meta-Analysis

Collectively, the results from Experiments 1–3 not only reveal early-emerging competence to reason about others' emotions, but also a difference between the two age groups. However, although we consistently found that 5-year-olds (but not 4-year-olds) differentiate the two characters' final emotional responses, each experiment was likely underpowered to identify a significant difference between age groups. In particular, post hoc power analysis revealed that our 5-year-old samples in Experiments 1 and 2 achieved 48% and 50% power, respectively. An internal meta-analysis can complement the limitations of insufficient power in individual experiments, and allow researchers to gain a more precise estimate of effect sizes (Braver, Thoemmes, & Rosenthal, 2014; Cumming, 2014). Given that we used the same task in all experiments with minimal variations in the procedure, we ran a mini meta-analysis (see Goh, Hall, & Rosenthal, 2016) across the three experiments, separately for each age group.

We fit a random-effects meta-analytic model using the *metafor* package in R (Viechtbauer, 2010), where the mean effect size in each experiment (Cohen's *d*; difference between emotion ratings for Low- and high-expectation characters) was weighted by sample size. Consistent with our findings, we found a highly significant effect for 5-year-olds ($\beta = .543$, $z = 3.102$, $p = .002$), but not 4-year-olds ($\beta = .010$, $z = .058$, $p = .954$). See Figure 5 for results. The results further corroborate our finding that although both 4- and 5-year-olds were able to appropriately raise or lower their emotion ratings depending on expectations (Experiment 1), only 5-year-olds make the adult-like inference that the low-expectation character feels better than the high-expectation character given identical outcomes.

Using the same meta-analytic approach, we also compared the magnitude of the main effect (difference in final emotion ratings between the two characters) between the two characters between 4-year-olds and 5-year-olds across all three experiments. The difference between age groups was highly significant, showing that the difference in ratings was larger in 5-year-olds than in 4-year-olds ($\beta = .477$, $z = 2.851$, $p = .004$).[5]

General Discussion

Across three experiments, we examined adults' and young children's ability to consider others' prior expectations to reason about their emotional responses to unexpected events. The main task involved inferring the emotional responses of two characters who had different expectations about future outcomes but experienced actual outcomes that were inconsistent with their expectations. Importantly, rather than providing explicit information about the characters' beliefs, we manipulated the course of physical events that would naturally lead participants to attribute different expectations to each character. Thus, the task capitalized on participants' naïve physics to provide contextual support for representing characters' beliefs about future outcomes; their expectations were plausibly grounded in how the bowling event might unfold.

In Experiment 1, 4-year-olds, 5-year-olds, and adults adjusted the characters' emotion ratings in the appropriate direction after seeing the final outcome, moving ratings upward for the low-expectation character and downward for the high-expectation character. Only the 5-year-old group made adjustments that were robust enough to result in higher final ratings for the low-expectation character than the high-expectation character. Experiments 2 and 3 suggested that 5-year-olds distinguished the agents' emotions without any explicit reference to their expectations, whereas 4-year-olds had difficulty even when prompted to think about the characters' expectations. Our meta-analysis further supported the difference between 4- and 5-year-old children in their ability to differentiate the two characters' emotional responses.

Together these results suggest that by five years of age, children already have an abstract, theory-like causal model of others' emotions: They understand how external events, actions, and mental states can together give rise to different affective states, and they can make systematic predictions about how others might feel based on available evidence (Gopnik & Wellman, 1992; Jara-Ettinger et al., 2016; Ong, Zaki, et al., 2015; Ong et al., 2018; Saxe & Houlihan, 2017; Wellman & Gelman, 1992). The current study adds to the growing body of work on children's understanding of the link between beliefs and emotions (e.g., Bradmetz & Schneider, 1999; Doan et al., 2018; Hadwin & Perner, 1991; Harris et al., 1989; Lagattuta, 2014; Lagattuta et al., 1997; Lara et al., 2017; Ronfard & Harris, 2014; Wellman & Liu, 2004; Wu & Schulz, 2018) and provides the earliest evidence for expectation-based emotion reasoning (see Doan et al., 2018; Lara et al., 2017). By age 4, children understand how others' emotional responses would change before and after the outcome depending on their prior expectations, and by age 5, children begin to appreciate that one agent might actually feel better than the other upon observing identical outcomes because of their prior expectations. Our results also show how basic cognitive capacities, such as physical reasoning abilities, can support rich, flexible understanding of others' emotional states, even when outcomes are insufficient and information about expectations are not explicitly communicated.

The meta-analysis of the final ratings summarizes a pattern that was consistent in all three studies: a rather striking difference between 4- and 5-year-old children in their ability to differentiate the two characters' emotional states. This pattern was robust to our different rating scales (verbal 6-point scale, physical 8-point scale, physical 6-point scale in Experiments 1–3, respectively); whether children were verbally responding or pointing at a scale, 5-year-olds showed a consistent differentiation between the two characters' emotions, and 4-year-olds did not. What drives the difference between the two age groups? Note that these results are consistent

---

[5] While the effect of age was clear between the age groups, it was not continuous. The correlation between age and the difference in ratings was weak, $r(138) = .142$, $p = .095$), and nearly absent in each age group, 4-year-olds: $r(68) = -.025$, $p = .835$; 5-year-olds: $r(68) = -.130$, $p = .283$.

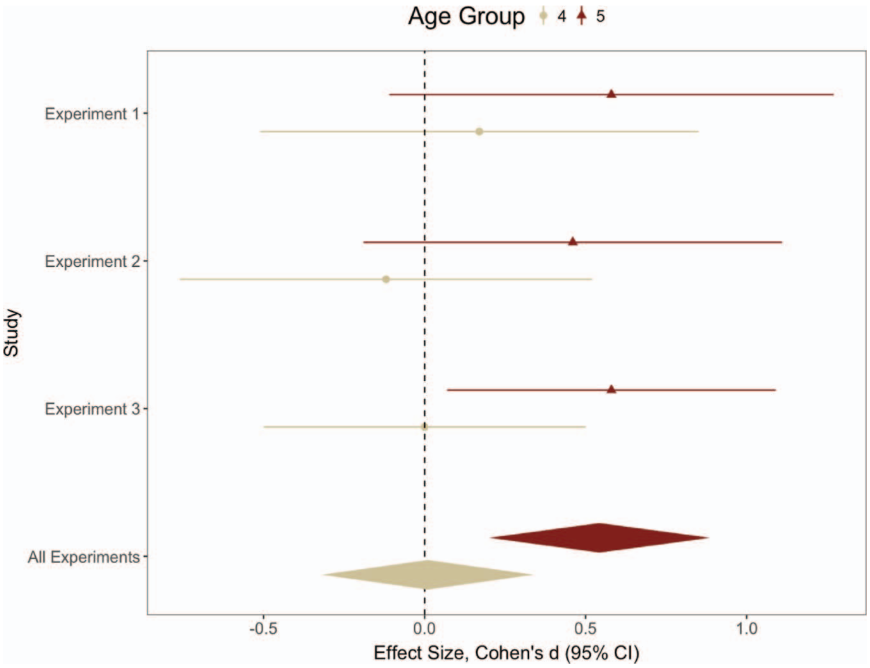

Figure 5. Forest plot showing effect size (Cohen's *d*) for Experiments 1–3 and the meta-analytic estimate of the effect size for each age group. Error bars represent 95% confidence intervals (CIs). See the online article for the color version of this figure.

with a recent study that also found a stark age difference during this period; while 5-year-olds readily infer both desires and beliefs based on changes in others' facial expressions, 4-year-olds have trouble inferring beliefs and can only infer desires (Wu & Schulz, 2018). However, our results further suggest that 4-year-olds' difficulty is not due to a total lack of competence to integrate expectations and outcomes to infer emotions. In Experiment 1, 4-year-olds appropriately adjusted their emotion ratings upward or downward depending on agents' prior expectations, even though these expections were no longer consistent with reality (cf. see De Rosnay et al., 2004; Harris et al., 1989), providing suggestive evidence that they have a basic understanding of how expectations modulate emotions. Their failure to differentiate the agents' final emotions, then, might reflect the possibility that these adjustments (especially the downward adjustment for the high-expectation character) were not robust enough to manifest as a clear difference in final ratings. We can entertain a few possible reasons here.

First, even though our rating measure is arguably more sensitive than binary forced-choice measure, it still requires careful training for preschool-aged children, and it is difficult to use with children younger than 4. In Experiment 2, we tried a simpler, verbal rating scale, and 4-year-olds still showed no difference between the two characters' emotions. Whether other measures (e.g., looking time, anticipatory looking) would better capture young children's abilities remains an open question for future work.

Second, it is possible that 4-year-olds had relatively weaker domain-specific knowledge to support their emotion inferences. Note for instance that success on our task requires the ability to generate rather fine-grained representations of the agents' expectations based on physical events (e.g., "Annie thinks that she will knock down six pins," or most pins, as opposed to any nonzero number of pins). Given that preschoolers' understanding of numbers larger than 3 (Sarnecka & Lee, 2009) and quantifiers (Horowitz, Schneider, & Frank, 2018) develops rapidly during this period, 4-year-olds might have generated less precise representations of the agents' expectations of the future physical outcomes compared to older children. This is consistent with our finding that although 4-year-olds could reliably predict *where* the ball could go, they had difficulty predicting the number of pins that would be knocked down, whereas 5-year-olds provided more accurate estimates (see online supplemental materials). Given weak, underspecified representations about the outcome, it might have been more challenging to estimate the discrepancy between expected and actual outcomes and subsequently to infer the character's emotional responses to the outcomes. The granularity of the underlying representations may also help explain why 5-year-olds judged the characters' emotions differently on the scale yet failed the binary judgment ("who feels better?"). Their ability to consider the discrepancy between agents' expectations and outcomes may be less robust compared to adults, and may only manifest in measures that are sensitive enough to reveal their understanding. Future work can investigate whether 5-year-olds can successfully make this binary judgment given more salient expectation-outcome discrepancies.

At a broader level, our results show how our real-world inferences about others' feelings are complex and often occur in conjunction with inferences in other domains beyond intuitive physics, such as number and probability. For instance, imagine one agent expects to hit one pin and the other expects to hit five pins, but both end up hitting a strike; would children understand that the former might feel happier than the latter? Success in such tasks may depend not only on children's ability to represent and reason about graded differences between the expected and actual out-

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

comes, but also inherently tied to their growing knowledge in numbers (Odic, Le Corre, & Halberda, 2015). Additionally, previous work with adults suggests that "near-misses" have a greater affective impact than "far-misses," resulting in more regret (Ong, Goodman, et al., 2015); that is, the relative probability of possible outcomes can influence our inferences about emotions. Our task matched the actual outcomes and thus did not manipulate whether the characters were near or far from completing their desired goal (e.g., strike); given recent work on school-age children's ability to infer surprise from probability information (Doan et al., 2018), investigating children's sensitivity to event probabilities in inferring different kinds of emotions may be a particularly promising line of research.

Here, we focused on an outcome (knocking down three pins) that was unexpected for both characters in our experiments. Prior work, however, has suggested that children evaluate others' emotions differently depending on the valence of the outcome (Lara et al., 2017). While a previous study with infants suggests that 10-month-olds understand the emotional consequences of positive outcomes (i.e., a completed goal before negative outcomes, Skerry & Spelke, 2014), another recent study with older children (Lara et al., 2017) found that children show earlier success in expectation-based emotion reasoning in scenarios that involve negative outcomes first (at 6–7 years) than attenuated outcomes (at 8–10 years), and failure in scenarios with positive outcomes even at age 10. In our task, both agents experienced an intermediate outcome that was not the best (hitting all pins) nor the worst (bowling the ball to the gutter), making the outcome most similar to the "attenuated outcome" in Lara et al. (2017). Given 5-year-olds' success in our task, it is possible that they would also understand that a character with high expectations feels worse than a character with low expectations when both experience a negative outcome. Testing children in scenarios that involve outcomes that are clearly positive or negative remains a goal for future work.

Children's developing understanding of others' skills or competence and how they relate to expected outcomes may play an important role in their reasoning about emotions. In our task, we intentionally provided participants with no information about the agents aside from the trajectory of their balls. However, as adults we understand that an expert bowler who only knocks down four pins might feel worse than a novice bowler who knocks down three pins on her first try. Adults also understand that sometimes skilled bowlers can make the ball change its trajectory even after it starts rolling (i.e., "hooking"); although knowledge of this skill was not relevant to the performance on this task, it is possible that participants considered these outcomes more or less surprising depending on whether they spontaneously brought these variables into account. Additionally, individual differences in traits (e.g., general optimism or pessimism) might also inform people's expectations about events and how they respond to unexpected outcomes (see Gibson & Sanbonmatsu, 2004). Our current results open up new questions about how we integrate agent-specific information (e.g., skill, knowledge, traits, habits) with objective physical information (e.g., the trajectory of the ball) to reason about others' expectations and emotions.

Our results in Experiment 1 show an interesting asymmetry in the amount of change in emotion ratings between the High- and low-expectation characters. For all age groups, the increase in happiness of the low-expectation character was larger than the amount of decrease in happiness of the high-expectation characters. This suggests that both children and adults might discount the negative impact of unexpected failures compared to the positive impact of unexpected successes. Although the current study does not address why participants in our task show this tendency, one possibility is that this effect may depend on the nature of the event; in the context of a fun game like bowling that does not involve significant investment of effort, positive outcomes can make us instantaneously "happy," while negative outcomes do not necessarily make us feel "sad." Furthermore, participants might have reasoned that the emotional significance of getting any reward when the agent expected to gain nothing is larger than the significance of gaining less than expected. A growing body of work suggests that humans reason about others' actions and mental states in light of their expected rewards and costs (naïve utility calculus; see Jara-Ettinger et al., 2016); while rewarding outcomes are generally considered to induce pleasure or happiness, further work is needed to understand how utility-based reasoning supports inferences about others' emotions.

So far, in describing the task and the results, we treated agents' beliefs as "prior" expectations, assuming that these are expectations that agents had in mind before the outcome was revealed. However, it is possible that children generated and considered agents' beliefs about what could have happened only after seeing the final outcome, especially in Experiment 2 where children were not probed about intermediate mental states and asked to infer their emotions after the final outcome has occurred. Although counterfactual reasoning is considered notoriously difficult for young children, some tasks suggest that even 5-year-olds can pass counterfactual thinking tasks (Beck, Robinson, Carroll, & Apperly, 2006; Nyhout & Ganea, 2019) and factor alternate outcomes into judgments of their own emotions (Weisberg & Beck, 2010). However, previous studies also suggest that counterfactual considerations do not influence children's judgments of others' counterfactual emotions such as regret or relief until age 7 (Ferrell, Guttentag, & Gredlein, 2009; Weisberg & Beck, 2010). The current study provides only suggestive evidence that 5-year-olds appreciate the role of counterfactuals in emotions; more compelling evidence could come from studies that directly manipulate how an agent's actions could have led to different outcomes, and how such decisions might generate feelings of regret or relief.

In sum, our results provide evidence for an early developing competence for expectation-based emotion reasoning as well as developmental change during preschool years. However, we also note a few limitations. First, our task was presented on a computer screen; even though we used animated movies to convey the dynamic nature of the physical event, the context was arguably artificial. Thus, it remains to be seen whether children's competence might manifest earlier in more ecologically valid contexts. Second, children in the current study were unable to make explicit binary judgments about who feels better of the two; the difference was found only in their relative ratings. Note that binary choices are quite noisy even in 7-year-olds (Doan et al., 2018). In fact, our measure is very similar to the one used in a closely relevant study (Lara et al., 2017); although children in the "attenuated outcome" condition (equivalent to our current test trials) did not differentiate their ratings until 8 years of age, we were able to identify a similar competence at a much younger age. While it would have been ideal to find converging evidence in both parametric and binary

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

measures, it is possible that the higher sensitivity of the rating measure captured underlying competence that would otherwise have gone unnoticed in a binary forced-choice question.

The ability to understand how others feel is an important aspect of our social intelligence. A body of developmental and computational work suggests that our ability to understand how others feel goes far beyond simple recognition of others' facial or bodily expressions, or statistical associations between emotions and outcomes; we have a rich, intuitive understanding of emotions that allows us to reason and explain how others felt in the past, predict what others will feel in the future, and even influence others' feelings by changing our own actions toward them. A hallmark of such abstract understanding of emotions is the ability to integrate expectations and outcomes. The current results provide the earliest evidence for expectation-based emotion reasoning in preschool-aged children, and offer some insights into what might change in development. Although it is easy to assume that developmental change in emotion-inference tasks are rooted in children's understanding of mental and affective states, it is important to consider how these inferences are often tied to children's growing knowledge in other domains. Beyond cross-domain inferences in physical causal reasoning (Schulz, Bonawitz, & Griffiths, 2007; Schulz & Gopnik, 2004), we hope our study will inspire more research on how children integrate and combine their intuitive theories across domains to reason about other minds (Ong et al., 2018). Such seamless integration may be one of the key secrets to the richness and the power of human social intelligence.

References

Baker, S. T., Leslie, A. M., Gallistel, C. R., & Hood, B. M. (2016). Bayesian change-point analysis reveals developmental change in a classic theory of mind task. *Cognitive Psychology, 91,* 124–149. http://dx.doi.org/10.1016/j.cogpsych.2016.08.001

Beck, S. R., Robinson, E. J., Carroll, D. J., & Apperly, I. A. (2006). Children's thinking about counterfactuals and future hypotheticals as possibilities. *Child Development, 77,* 413–426. http://dx.doi.org/10.1111/j.1467-8624.2006.00879.x

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. http://dx.doi.org/10.1348/026151099165438

Braver, S. L., Thoemmes, F. J., & Rosenthal, R. (2014). Continuously cumulating meta-analysis and replicability. *Perspectives on Psychological Science, 9,* 333–342. http://dx.doi.org/10.1177/1745691614529796

Carey, S. (1985). *Conceptual change in childhood*. Cambridge, MA: MIT Press.

Cumming, G. (2014). The new statistics: Why and how. *Psychological Science, 25,* 7–29. http://dx.doi.org/10.1177/0956797613504966

Cushman, F., Sheketoff, R., Wharton, S., & Carey, S. (2013). The development of intent-based moral judgment. *Cognition, 127,* 6–21. http://dx.doi.org/10.1016/j.cognition.2012.11.008

Denison, S., Trikutam, P., & Xu, F. (2014). Probability versus representativeness in infancy: Can infants use naïve physics to adjust population base rates in probabilistic inference? *Developmental Psychology, 50,* 2009–2019. http://dx.doi.org/10.1037/a0037158

De Rosnay, M., Pons, F., Harris, P. L., & Morrell, J. M. (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. http://dx.doi.org/10.1348/026151004323044573

Doan, T., Friedman, O., & Denison, S. (2018). Beyond belief: The probability-based notion of surprise in children. *Emotion, 18,* 1163–1173. http://dx.doi.org/10.1037/emo0000394

Ferrell, J. M., Guttentag, R. E., & Gredlein, J. M. (2009). Children's understanding of counterfactual emotions: Age differences, individual differences, and the effects of counterfactual-information salience. *British Journal of Developmental Psychology, 27,* 569–585. http://dx.doi.org/10.1348/026151008X337743

Filippova, E., & Astington, J. W. (2008). Further development in social reasoning revealed in discourse irony understanding. *Child Development, 79,* 126–138. http://dx.doi.org/10.1111/j.1467-8624.2007.01115.x

Flavell, J. H. (1999). Cognitive development: Children's knowledge about the mind. *Annual Review of Psychology, 50,* 21–45. http://dx.doi.org/10.1146/annurev.psych.50.1.21

Gergely, G., & Csibra, G. (2003). Teleological reasoning in infancy: The naïve theory of rational action. *Trends in Cognitive Sciences, 7,* 287–292. http://dx.doi.org/10.1016/S1364-6613(03)00128-1

Gibson, B., & Sanbonmatsu, D. M. (2004). Optimism, pessimism, and gambling: The downside of optimism. *Personality and Social Psychology Bulletin, 30,* 149–160. http://dx.doi.org/10.1177/0146167203259929

Gnepp, J., & Hess, D. L. (1986). Children's understanding of verbal and facial display rules. *Developmental Psychology, 22,* 103–108. http://dx.doi.org/10.1037/0012-1649.22.1.103

Goh, J. X., Hall, J. A., & Rosenthal, R. (2016). Mini meta-analysis of your own studies: Some arguments on why and a primer on how. *Social and Personality Psychology Compass, 10,* 535–549. http://dx.doi.org/10.1111/spc3.12267

Gopnik, A., Sobel, D. M., Schulz, L. E., & Glymour, C. (2001). Causal learning mechanisms in very young children: Two-, three-, and four-year-olds infer causal relations from patterns of variation and covariation. *Developmental Psychology, 37,* 620–629. http://dx.doi.org/10.1037/0012-1649.37.5.620

Gopnik, A., & Wellman, H. M. (1992). Why the child's theory of mind really is a theory. *Mind & Language, 7,* 145–171. http://dx.doi.org/10.1111/j.1468-0017.1992.tb00202.x

Gweon, H., Dodell-Feder, D., Bedny, M., & Saxe, R. (2012). Theory of mind performance in children correlates with functional specialization of a brain region for thinking about thoughts. *Child Development, 83,* 1853–1868. http://dx.doi.org/10.1111/j.1467-8624.2012.01829.x

Gweon, H., Tenenbaum, J. B., & Schulz, L. E. (2010). Infants consider both the sample and the sampling process in inductive generalization. *Proceedings of the National Academy of Sciences of the United States of America, 107,* 9066–9071. http://dx.doi.org/10.1073/pnas.1003095107

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

Harris, P. L. (1989). *Children and emotion: The development of psychological understanding*. Cambridge, MA: Basil Blackwell.

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

Horowitz, A. C., Schneider, R. M., & Frank, M. C. (2018). The trouble with quantifiers: Exploring children's deficits in scalar implicature. *Child Development, 89,* e572–e593. http://dx.doi.org/10.1111/cdev.13014

Jara-Ettinger, J., Gweon, H., Schulz, L. E., & Tenenbaum, J. B. (2016). The naïve utility calculus: Computational principles underlying commonsense psychology. *Trends in Cognitive Sciences, 20,* 589–604. http://dx.doi.org/10.1016/j.tics.2016.05.011

Kotovsky, L., & Baillargeon, R. (2000). Reasoning about collisions involving inert objects in 7.5-month-old infants. *Developmental Science, 3,* 344–359. http://dx.doi.org/10.1111/1467-7687.00129

Lackner, C. L., Bowman, L. C., & Sabbagh, M. A. (2010). Dopaminergic functioning and preschoolers' theory of mind. *Neuropsychologia, 48,* 1767–1774. http://dx.doi.org/10.1016/j.neuropsychologia.2010.02.027

Lagattuta, K. H. (2014). Linking past, present, and future: Children's ability to connect mental states and emotions across time. *Child Development Perspectives, 8,* 90–95. http://dx.doi.org/10.1111/cdep.12065

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. http://dx.doi.org/10.1111/j.1467-8624.1997.tb01986.x

Lara, K. H., Lagattuta, K. H., & Kramer, H. J. (2017). Is there a downside to anticipating the upside? Children's and adults' reasoning about how prior expectations shape future emotions. *Child Development*. Advance online publication. http://dx.doi.org/10.1111/cdev.12994

Mellers, B. A., Schwartz, A., Ho, K., & Ritov, I. (1997). Decision affect theory: Emotional reactions to the outcomes of risky options. *Psychological Science, 8,* 423–429. http://dx.doi.org/10.1111/j.1467-9280.1997.tb00455.x

Muentener, P., & Carey, S. (2010). Infants' causal representations of state change events. *Cognitive Psychology, 61,* 63–86. http://dx.doi.org/10.1016/j.cogpsych.2010.02.001

Newman, G. E., Keil, F. C., Kuhlmeier, V. A., & Wynn, K. (2010). Early understandings of the link between agents and order. *Proceedings of the National Academy of Sciences of the United States of America, 107,* 17140–17145. http://dx.doi.org/10.1073/pnas.0914056107

Nyhout, A., & Ganea, P. A. (2019). Mature counterfactual reasoning in 4- and 5-year-olds. *Cognition, 183,* 57–66. http://dx.doi.org/10.1016/j.cognition.2018.10.027

Odic, D., Le Corre, M., & Halberda, J. (2015, May). Children's mappings between number words and the approximate number system. *Cognition, 138,* 102–121. http://dx.doi.org/10.1016/j.cognition.2015.01.008

Ong, D. C., Goodman, N. D., & Zaki, J. (2015). Near-misses sting even when they are uncontrollable. In D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, & P. P. Maglio (Eds.), *Proceedings of the thirty-seventh annual meeting of the Cognitive Science Society* (pp. 1775–1780). Austin, TX: Cognitive Science Society.

Ong, D. C., Zaki, J., & Goodman, N. D. (2015). Affective cognition: Exploring lay theories of emotion. *Cognition, 143,* 141–162. http://dx.doi.org/10.1016/j.cognition.2015.06.010

Ong, D. C., Zaki, J., & Goodman, N. D. (2018). Computational models of emotion inference in theory of mind: A review and roadmap. *Topics in Cognitive Science*. Advance online publication. http://dx.doi.org/10.1111/tops.12371

Reschke, P. J., Walle, E. A., Flom, R., & Guenther, D. (2017). Twelve-month-old infants' sensitivity to others' emotions following positive and negative events. *Infancy, 22,* 874–881. http://dx.doi.org/10.1111/infa.12193

Ronfard, S., & Harris, P. L. (2014, January). When will Little Red Riding Hood become scared? Children's attribution of mental states to a story character. *Developmental Psychology, 50,* 283–292. http://dx.doi.org/10.1037/a0032970

Sarnecka, B. W., & Lee, M. D. (2009). Levels of number knowledge during early childhood. *Journal of Experimental Child Psychology, 103,* 325–337. http://dx.doi.org/10.1016/j.jecp.2009.02.007

Saxe, R., & Houlihan, S. D. (2017). Formalizing emotion concepts within a Bayesian model of theory of mind. *Current Opinion in Psychology, 17,* 15–21. http://dx.doi.org/10.1016/j.copsyc.2017.04.019

Saxe, R., Tenenbaum, J. B., & Carey, S. (2005). Secret agents: Inferences about hidden causes by 10- and 12-month-old infants. *Psychological Science, 16,* 995–1001. http://dx.doi.org/10.1111/j.1467-9280.2005.01649.x

Schulz, L. E., Bonawitz, E. B., & Griffiths, T. L. (2007). Can being scared cause tummy aches? Naive theories, ambiguous evidence, and preschoolers' causal inferences. *Developmental Psychology, 43,* 1124–1139. http://dx.doi.org/10.1037/0012-1649.43.5.1124

Schulz, L. E., & Gopnik, A. (2004). Causal learning across domains. *Developmental Psychology, 40,* 162–176. http://dx.doi.org/10.1037/0012-1649.40.2.162

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

Spelke, E. S., Breinlinger, K., Macomber, J., & Jacobson, K. (1992). Origins of knowledge. *Psychological Review, 99,* 605–632. http://dx.doi.org/10.1037/0033-295X.99.4.605

Spelke, E. S., Katz, G., Purcell, S. E., Ehrlich, S. M., & Breinlinger, K. (1994). Early knowledge of object motion: Continuity and inertia. *Cognition, 51,* 131–176. http://dx.doi.org/10.1016/0010-0277(94)90013-2

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36,* 1–48. http://dx.doi.org/10.18637/jss.v036.i03

Weisberg, D. P., & Beck, S. R. (2010). Children's thinking about their own and others' regret and relief. *Journal of Experimental Child Psychology, 106,* 184–191. http://dx.doi.org/10.1016/j.jecp.2010.02.005

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. http://dx.doi.org/10.1111/j.2044-835X.1991.tb00871.x

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

Wellman, H. M., & Gelman, S. A. (1992). Cognitive development: Foundational theories of core domains. *Annual Review of Psychology, 43,* 337–375. http://dx.doi.org/10.1146/annurev.ps.43.020192.002005

Wellman, H. M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Development, 75,* 523–541. http://dx.doi.org/10.1111/j.1467-8624.2004.00691.x

Wellman, H. M., & Woolley, J. D. (1990). From simple desires to ordinary beliefs: The early development of everyday psychology. *Cognition, 35,* 245–275. http://dx.doi.org/10.1016/0010-0277(90)90024-E

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

Wu, Y., Baker, C. L., Tenenbaum, J. B., & Schulz, L. E. (2018). Rational inference of beliefs and desires from emotional expressions. *Cognitive Science, 42,* 850–884. http://dx.doi.org/10.1111/cogs.12548

Wu, Y., Muentener, P., & Schulz, L. E. (2017). One- to four-year-olds connect diverse positive emotional vocalizations to their probable causes. *Proceedings of the National Academy of Sciences of the United States of America, 114,* 11896–11901. http://dx.doi.org/10.1073/pnas.1707715114

Wu, Y., & Schulz, L. E. (2018). Inferring beliefs and desires from emotional reactions to anticipated and observed events. *Child Development, 89,* 649–662. http://dx.doi.org/10.1111/cdev.12759

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. http://dx.doi.org/10.1111/j.2044-835X.1984.tb00536.x

Received November 6, 2018
Revision received March 29, 2019
Accepted April 2, 2019 ■

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.