Young children and adults integrate past expectations and current outcomes to reason about others' emotions

Desmond C. Ong[1] **(dco@stanford.edu)** **Mika Asaba**[1] **(masaba@stanford.edu)**
Hyowon Gweon (hyo@stanford.edu)
Department of Psychology, Stanford University, Stanford CA, USA

Abstract

Reasoning about others' emotions is a crucial component in social cognition. Here, we tested the ability of preschool children to reason about an agent's emotions following an unexpected outcome. Importantly, we controlled for the actual payoff of the outcome, while varying the prior expectation of the agents. Five-year-olds, but not four-year-olds, were able to correctly judge an agent's emotions following an unexpected outcome (Experiment 1). When explicitly provided with the agent's expectations, 4-year-olds were then also able to correctly judge the agent's feelings (Experiment 2). Our results suggest that the ability to reason about emotions given outcomes and prior expectations develops by 4 years of age, while the ability to spontaneously infer such prior expectations develops soon after. We discuss our results in light of the developmental literature on emotion understanding and counterfactual reasoning.

Keywords: Affective Reasoning; Affective Cognition; Theory of Mind; counterfactual reasoning

As much as the ability to feel is a fundamental part of what makes us human, the ability to *reason* about how *others* feel—what we call Affective Cognition (Ong, Zaki, & Goodman, 2015)—is also core to who we are. We have a rich and 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 intervene by changing our own actions to influence how others will feel.

As we reason about others' emotions, we often have to rely on unobservable mental states that are internal to agents (e.g., goals, beliefs, desires), especially in the absence of explicit cues (e.g., facial expressions or the valence of events). Previous developmental studies have found surprisingly early signatures of understanding the relationship between people's mental states, goal-directed actions, event outcomes, and affective states. For instance, infants can differentiate between actions that are congruent or incongruent with the affect that agents previously expressed towards objects (Phillips, Wellman, & Spelke, 2002). Furthermore, infants distinguish between displayed emotion that is congruent or incongruent with the outcomes of agent's goal-directed actions (Skerry & Spelke, 2014). For instance, 10-month-olds showed more surprise when an agent successfully achieved its goal (e.g., jumping over a barrier) and expressed incongruent (negative) affect, as compared to when the agent displayed congruent (positive) affect. However, when the agent failed to achieve its goal, infants did not distinguish between congruent (negative) and incongruent (positive) affect (Skerry & Spelke, 2014). By preschool, children can also understand that failed outcomes lead to negative affect (e.g., Barden, Zelko, Duncan, & Masters, 1980; Wellman & Banerjee, 1991; Wu, Baker, Tenenbaum, & Schulz, 2014).

However, such competence might not necessarily suggest that children have a full-fledged, theory-like understanding of others' goals, actions, and affective states by preschool years. For instance, children's ability to link successes with positive emotions (and failures with negative emotions) might reflect mere statistical associations that young children have acquired from early social experience. As adults, we intuitively understand that the same outcome can have very different affective consequences depending on the agent's prior beliefs, or expectations, about what would happen. Importantly, these expectations cannot be explicitly observed from others' actions, but must be generated by reasoning about the agent's mental states, the structure of the physical world, and the agent's interactions with the physical world. In this study, we focus on children's ability to integrate an agent's prior expectation with an outcome when predicting their affect.

Imagine this scene at the bowling alley. Sally throws her ball: it heads towards the gutter but remarkably, curves back to finally knock down three pins. Annie's ball rolls straight towards the center pin, but it curves to the left and only hits three pins. Although Sally and Annie experienced identical, unexpected outcomes (i.e., hitting three pins), one might reasonably infer that Sally feels happier than Annie. Although seemingly intuitive, this conclusion requires a complex set of inferences, involving our understanding of the physical and social world. To judge Sally as happier than Annie, we must (1) understand where each ball *could* go and project this to our friend's expectations, (2) link these expectations with affective states, (3) calculate the discrepancy between the expected and the actual outcome, and (4) judge each character's emotional responses with respect to the expected and actual outcome. An adult-like prediction of the characters' emotions (e.g., Annie as disappointed and Sally, relieved) involves integrating information from various domains.

Decades of developmental research have shown how each of these components independently undergo their own developmental trajectory in early childhood. First, understanding where the ball could go requires an understanding of how objects move and how they respond to causal interventions from agents. It takes years to construct a coherent, intuitive theory of physics that supports accurate inferences and goal-directed actions from infancy (Spelke, Breinlinger, Macomber, & Jacobson, 1992) to the preschool years (Hood, Cole-Davies, & Dias, 2003; Keen, 2003). Second, children also show both surprising competence as well as a significant developmental

[1]These authors contributed equally to this work.

change in their understanding of others' emotions as reflected in their display cues (e.g. Gnepp & Hess, 1986) and consequences of their motives and behaviors (e.g. Yuill, 1984). More broadly, young children undergo significant changes in using others' mental states to predict and explain behaviors (Baillargeon, Scott, & He, 2010; Gweon, Dodell-Feder, Bedny, & Saxe, 2012; Wellman, Cross, & Watson, 2001). It is an open question as to how children integrate inferences of others' mental states (e.g., goals and beliefs) with knowledge about the physical world to reason about affective states.

Across two studies, we examine whether preschool children utilize others' prior expectations (i.e., others' beliefs about what they expected to happen) to draw different conclusions about an agent's affective response to identical physical outcomes. Specifically, we adapted the bowling scenario introduced above into a simple task. In Experiment 1, we examined children and adults' inferences about others' emotions after having observed the full scenario. In Experiment 2, we break down the task into two parts and explicitly provide the character's expectations to uncover a more fine-grained understanding of the inferences involved in the final judgment.

Experiment 1

In Experiment 1[2], we examined how children and adults reason about agents' emotions following an unexpected outcome. Participants observed characters bowling and provided emotion ratings for each character.

Participants. Children: Twenty 4-year-olds (8 female, M_{Age}(SD): 4.53(.26), range: 4.19 - 4.98) and twenty 5-year-olds (13 female, M_{Age}(SD): 5.32(.16), range: 5.06 - 5.61) were recruited from a campus preschool and local museum. One additional child was tested and dropped from analysis due to failure to respond to test questions. Adults: Fifty nine adults (17 female, M_{Age}(SD): 32.8 (11.2), range: 20 - 68) were recruited through Amazon's Mechanical Turk (AMT). One additional adult participated but was dropped for providing an incorrect answer to a planned attention check question ("How many pins were knocked down?").

Materials. Children first played a warm-up game using a child-friendly toy bowling set: six colorful foam pins, a red foam ball, and a black tarp for the bowling lane. We ran the main experiment on a computer using cartoons animated in Keynote. The cartoon bowling alley consists of a wide beige rectangle flanked by two narrow grey rectangles (the "gutter"; see Fig. 1). We designed the cartoon pins and balls to match the bowling set that children had played with earlier. We used simple, generic cartoon characters for the practice trials, and Elmo and Cookie Monster for the test trials. In all trials, the characters' backs were facing the child such that no facial expressions were shown. We used the same stimuli with adults.

[2]All experiments, data, and analyses can be found at: https://github.com/desmond-ong/bowling

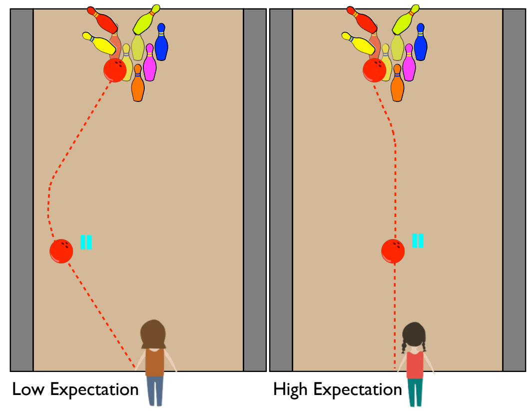

Figure 1: Sample stimuli used for Exp. 1 and 2. Left: Low Expectation video; Right: High Expectation video. Trajectories of the bowling balls are indicated by the dashed red lines. For Exp. 2, the position of the bowling balls at the midway pause are given next to the cyan "Pause" icons.

Procedures. Child Experiment: Children were tested in a quiet room inside of a campus preschool or local museum. The experimenter introduced the game of bowling by explaining that the goal is to use the ball to knock down as many pins as possible; children played with the bowling set for a few minutes. The experimenter then introduced the virtual bowling game on the computer. Children were asked how many pins were on the screen ("six") and what it meant when the bowling ball entered the gray areas ("it's out"). If a child could not count correctly or did not remember the answer, the experimenter repeated the explanation. Before the main experiment, all children were able to correctly report the number of pins and the meaning of the gray area.

There were four trial types: two practice trials presented in a fixed order (Gutter, Strike), and two test trials in a counterbalanced order (Low Expectation, High Expectation). In the Gutter trial, the character rolled the ball, which curved to the right, into the gray area, knocking down no pins. In the Strike trial, the character's ball went straight and knocked down all six pins. After each trial, the experimenter asked the child how many pins were knocked down and whether the character is feeling happy or sad. If the child said happy (sad), the experimenter asked, "kind of happy (sad), medium happy (sad), or really happy (sad)?" We converted children's verbal ratings to a 6-point scale, from "really sad" (1) to "really happy" (6), to be used as our Final Rating. After the practice trials, the experimenter asked: "Who do you think is happier?" (Forced Choice). If the child did not respond or said both, the experimenter asked the child to select one character.

The test trials had an identical sequence of prompts. In the Low Expectation trial, the ball initially headed sharply left towards the gutter but then curved back to hit three pins. In the High Expectation trial, the ball initially headed straight down the lane but then curved to the left to hit three pins. We counterbalanced the order and characters for the test trials.

Adult Experiment: Adult participants participated in a similar experiment with only the Low and High Expectation trials (no practice trials). After each trial, they providing an emotion rating for the character on a 6 point Likert scale from "really sad" (1) to "really happy" (6). After both trials, participants were asked to choose who is happier; adults also had the option of reporting that the characters are equally happy.

Results. Children: All children correctly reported the number of pins knocked down in every trial. There were no significant effects of trial order or characters. For the practice trials, 4- and 5-year-olds provided higher ratings to the Strike character than the Gutter character (4 year-olds: Strike vs. Gutter: M(SD) = 5.95(.22) vs. 1.30(.57), $t(19) = 35.42$, $p < 0.001$; 5 year-olds: Strike vs. Gutter: M(SD) = 5.95(.22) vs. 1.70(.57), $t(19) = 29.76$, $p < 0.001$). When given a binary choice between the Strike and Gutter character, all children in both age groups reported that the Strike character was happier than the Gutter character (4-year-olds: 20 of 20, $p < 0.001$; 5-year-olds: 20 of 20, $p < 0.001$).

Next, we examined ratings for the test trials (Fig. 2). Recall that both trials have identical outcomes (3 pins knocked down) but they differ with respect to the balls' initial trajectories (and thus agents' prior expectations). Four year-olds' provided the same ratings for the Low Expectation and High Expectation characters (Low Exp. vs. High Exp.: M(SD) = 4.70(1.34) vs. 4.85(1.14), $t(19) = -.55$, $p = ns$), whereas five year-olds reported higher ratings for the Low Expectation character than for the High Expectation character (Low Exp. vs. High Exp.: M(SD) = 4.80(.95) vs. 4.30(1.22), $t(19) = 2.13$, $p = .046$). When given a forced choice between the Low Expectation and High Expectation character, neither group showed a strong tendency to choose the Low Expectation character (binomial test, $p = ns$ for both), but we note that more five-olds chose this character over the High Expectation character (13 of 20 5-year-olds vs. 9 of 20 4-year-olds).

Adults: Adults' responses were consistent with the 5 year-olds' ratings, providing higher emotion ratings to the Low Expectation character than the High Expectation character (Low Exp. vs. High Exp.: M(SD) = 4.39 (.74) vs. 3.86 (.80), $t(58) = 4.60$, $p < 0.001$). On the final forced choice question, 23 (of 59) participants chose the Low Expectation character as feeling happier, 2 chose the High Expectation character, and the remaining 34 said that both feel equally happy.

Our results revealed a difference between the age groups: Both adults and 5-year-olds provided more positive emotion ratings to the Low Expectation character and lower ratings to the High Expectation character. Four-year-olds were not able to differentiate between these trials. We consider two possibilities for 4-year-olds' failures. First, 4-year-olds might understand the physical trajectories of the ball and how it relates to the characters' expectations, but have difficulties inferring the characters' emotional responses with respect to the discrepancies between expectations and outcomes.

Another possibility is that 4-year-olds have difficulty using their physical understanding to spontaneously generate the characters' expectations. Additional pilot data suggests that 4-year-olds indeed have a weak understanding of the physics

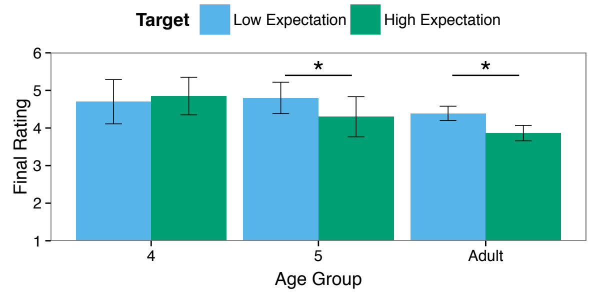

Figure 2: Exp. 1 Results. Participants' ratings of the characters' happiness after the test trials. Error bars indicate 95% confidence intervals. * indicates $p < .05$

involved in the experiment. We showed fourteen 4-year-olds (M_{Age}(SD): 4.66(.32), range: 4.12 - 4.99) the same videos, but with a pause in the middle. We asked children to predict where the ball would go and how many pins would be knocked down for the Low Expectation and High Expectation characters. Indeed, we found that children had great difficulty accurately predicting the final trajectory ("out" or "straight") and number of pins for the Low Expectation and High Expectation characters: only 7 of 14 children provided accurate responses for both characters. Thus, 4-year-olds' difficulty on this task may stem from a genuine inability to spontaneously generate the alternative outcome (where the ball is going to go and how many pins will be knocked down), which impairs their ability to utilize this alternative to infer the agent's emotion following the actual outcome.

Experiment 2

Experiment 2 tests the hypothesis that given explicit information about the agents' expectations, even 4-year-olds may be able to distinguish the two characters' emotions despite the same final outcomes. We used the videos from Exp. 1, but paused the videos halfway and explicitly provided participants with the characters' expectations about the ball's trajectory and how many pins will be knocked down.

Participants. Children. Eighteen 4-year-olds (8 female, M_{Age}(SD): 4.51(.32), range: 4.00 - 4.97), and eighteen 5-year-olds (9 female, M_{Age}(SD): 5.43(.27), range: 5.05 - 5.98), were recruited from a preschool and local museum. An additional 7 4-year-olds and 3 5-year-olds were tested and dropped due to incorrect responses to the initial emotion question (i.e., reporting the Low Expectation character to be happy or the High Expectation character to be sad; see Procedure). Adults. Fifty-eight adult participants were recruited through AMT (32 female; M_{Age}(SD) = 32.8(10.4), range: 18-64). An additional two participants were tested and dropped after failing planned check questions.

Materials. The stimuli were similar to those in Experiment 1, but we used the generic cartoon characters for the Low Expectation and High Expectation trials. The critical change was that the videos in the Low Expectation and High Expectation trials paused when the ball was halfway down the alley, and the experimenter explicitly provided the characters' expectations. To measure children's inferences about changes in emotion, we constructed a 10″ x 2″ physical scale with large arrows at each end. The scale was divided into eight equal sections with a large red sad face on the third section and a large green happy face on the fifth section (see Fig. 3).

Procedures. Child experiment: The procedures were similar to that of Experiment 1. After the warm-up game, children sequentially observed the Gutter and Strike practice trials and reported the number of pins that were knocked down. All children correctly reported the final outcomes. The experimenter then asked whether the character was feeling happy or sad (binary choice); children responded by putting a small magnet versions of the characters onto either the happy or sad face on the scale. After the Gutter and Strike trials, the experimenter showed both characters on the screen and asked the child to choose who is happier (forced choice).

Children then saw the Low Expectation and High Expectation test trials in a counterbalanced order. Halfway through the video, the video paused; this location is indicated by the ball and the cyan "Pause" icon in Figure 1 (the icon did not show up in the actual videos). The experimenter provided the character's belief about the ball's trajectory and the number of pins the ball would hit. In the Low Expectation trial, the video was paused when the ball was close to the gutter on the left side. The experimenter said, "Sally thinks that her ball is going to *go out* and hit *none* of the pins!" Children then rated how Sally feels right now ("happy or sad") by placing a Sally-shaped magnet on the scale (Initial Rating). When the video resumed, Sally's ball curved back and hit three pins. As an attention check, children were first asked how many pins were knocked down ("3"). The experimenter then reminded the child of their earlier emotion rating, and asked for a judgment of change, e.g., "Earlier you said she was feeling 'happy'. Do you think she is feeling better, or worse now?" (Verbal Report). The experimenter prompted 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 was used as the Final Rating.

The trial structure was identical in the High Expectation trial, except that the video was paused with the ball in the middle of the lane. The experimenter said, "Annie thinks that her ball is going to *go straight* and hit *all* of the pins!" After the child provided an Initial Rating by placing the Annie magnet piece on the scale, the video resumed to show Annie's ball continuing straight down the alley before curving left to knock down three pins. Children then answered how many pins she knocked down (Attention Check), indicated whether she feels better (Verbal Report), and provided a final emotion rating by moving the magnet piece on the scale (Final Rating). Last, after both test trials, the experimenter asked, "Who do you think is happier? Sally or Annie?" (Forced Choice). If the child did not respond or said both, the experimenter prompted the child to choose just one character.

Adult experiment: Adult participants participated in a similar experiment, with a few small changes. Adult participants were given the character's belief about the trajectory ("Annie thinks her ball is going straight / going out") and were prompted to predict how many pins the character expects to knock down (free-response from 0 to 6). Participants provided an initial happy or sad rating, observed the rest of the video, indicated whether the character feels better or worse, and provided a final rating for the character. Finally, after both trials, adult participants were asked to choose who is happier; they had the option of saying both are equally happy.

Results. All children correctly answered the number of pins knocked down in each trial. We observed no effect of trial order or character. For the Strike and Gutter practice trials, both 4- and 5-year-olds successfully reported that the Gutter character was sad (4 year-olds: 17 of 18; 5 year-olds: 18 of 18, $p < 0.001$ by binom. test for both) and the Strike character was happy (4 year-olds: 17 of 18; 5 year-olds: 17 of 18, $p < 0.001$ by binom. test for both). All children in both age groups chose the Strike character as happier than the Gutter character (4-year-olds: 18 of 18; 5-year-olds: 18 of 18, $p < 0.001$ by binom. test for both).

For the critical Low Expectation and High Expectation trials, we analyzed verbal responses ("better or worse?"), relative ratings (difference between the Initial Ratings and Final Ratings), and Final Ratings for each trial. Only children who correctly reported that the Low Expectation character was sad and the High Expectation character was happy in the Initial Ratings were included in our analyses. The vast majority of children reported that the Low Expectation character felt better after the outcome than at the pause (4 year olds: 16 of 18, $p = .001$; 5 year-olds: 17 of 18, $p < 0.001$), while children were at chance for reporting the High Expectation character to be feeling worse (4 year olds: 12 of 18; 5 year olds: 10 of 18). However, 5-year-olds showed a significant difference in their better/worse responses for the characters ($p = .02$, Fisher's exact) and 4-year-olds showed a marginal difference ($p = .09$). Next, we looked at children's relative ratings. Both age groups moved the Low Expectation character higher on the scale and the High Expectation character lower (4 year-olds: Low Exp. vs. High Exp.: M(SD) = 2.44(1.21) vs. -.83(1.91), $t(17) = -5.39$, $p < 0.001$; 5 year-olds: Low Exp. vs. High Exp.: M(SD) = 2.53(1.8) vs. -.86(2.39), $t(17) = -8.06$, $p < 0.001$). We found that 4-year-olds' final ratings for the Low Expectation and High Expectation characters did not differ (Low Exp. vs. High Exp.: M(SD) = 5.17(1.91) vs. 5.44(1.21), $t(17) = .79$, $p = ns$), whereas 5-year-olds rated the Low Expectation character as being happier than the High Expectation character (Low Exp. vs. High Exp.: M(SD) = 5.86(1.66) vs. 4.81(2.11), $t(17) = 2.60$, $p = .05$). Lastly, given a binary forced choice, neither group was able to report that the Low Expectation character was happier than the High Expectation character ($p = ns$ for both). However, 5-year-olds tended to choose the Low Expectation more often (12 of 18 5-year-olds vs. 8 of 18 4-year-olds).

Overall, adult participants responded as we predicted on

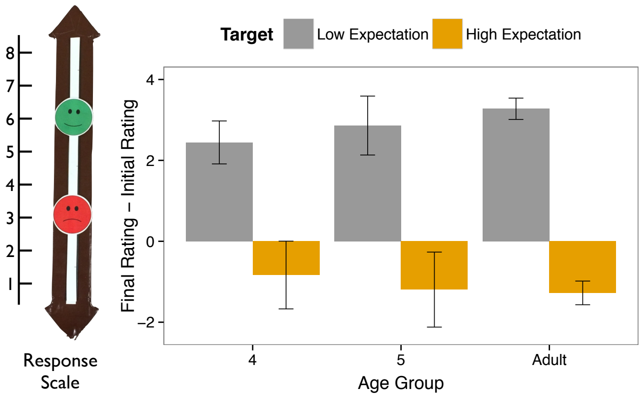

Figure 3: Left: Physical scale used in Exp. 2. Right: Participants' change in rating (before and after the outcome). Error bars represent 95% CIs.

all the questions. Participants reported that the Low Expectation character felt sad initially (57 of 58) and felt better after seeing the outcome (58 of 58); they thought the High Expectation character felt happy initially (56 of 58) and felt worse after the outcome (46 of 58). Adults' ratings were significantly higher for the Low Expectation character (Low Exp. vs. High Exp.: M(SD) = 6.33 (.98) vs. 4.62 (0.89), $t(57)=10.46$, $p<0.001$). On the forced choice question, 42 chose the Low Expectation character as feeling happier, and the remaining 16 said that both characters feel equally happy.

Given explicit information about characters' expectations, four- and five-year-olds showed an adult-like response, shifting their ratings in the appropriate directions after observing the final outcomes. Five-year-olds and adults, but not four-year-olds, providing higher final ratings for the Low Expectation character than the High Expectation character. Four-year-olds' relative differences between the Initial and Final ratings were not as robust as those of five-year-olds and adults, resulting in this lack of difference in the Final Rating score. One possibility is that younger children were simply rating the characters' emotions with respect to the final outcome (e.g., how each character feels after knocking down 3 pins), explaining the change in rating and the lack of difference between final ratings. However, 4-year-olds' "Better" and "Worse" verbal reports suggest that they are differentiating between these characters' emotions based on prior beliefs.

General Discussion

Across two experiments, we examined how children use an agent's prior expectations to judge how they would feel following identical, unexpected outcomes. In Exp. 1, We found a developmental shift between ages 4 - 5 in children's inferences about others' emotions. Given two characters with identical, unexpected outcomes, older children gave adult-like responses; a character who had a low expectation about the outcome would feel better than a character who had a high expectation. However, younger children (4-year-olds) were unable to distinguish the two characters. In Exp. 2, given explicit information about the characters' expectations, however, even the younger children appropriately adjusted their ratings about the character's feelings. Our results suggest that young children can already make use of an agent's prior expectations to reason about their relative emotions, but the ability to spontaneously infer those prior expectations continue to develop in preschool years.

Our study examined judgments about current emotions based on expectations, but we also often base these judgments from alternative states or *what could have happened* (contrary to what actually happened), a capacity often referred to as counterfactual reasoning. Five-year-olds can pass counterfactual thinking tasks (e.g., Beck, Robinson, Carroll, & Apperly, 2006) and factor alternate outcomes into judgments of their own emotions (Weisberg & Beck, 2010, 2012). However, previous studies suggest that counterfactual considerations do not factor into children's judgments of others' counterfactual emotions (*regret* and *relief*) until later in development, at 7 years of age (Ferrell, Guttentag, & Gredlein, 2009; Weisberg & Beck, 2010). Our findings provide evidence for a precursory capacity to reasoning about counterfactual emotions. Our next step is to more directly address children's understanding of alternative states and counterfactual emotions using a similar paradigm as the one used here.

An additional question to consider is exactly how and when children generate relevant alternative states. In our task, children's ability to reason about the physical states of the world was an important actor for generating these alternative states. Even though infants can make sophisticated predictions about physical outcomes (Spelke et al., 1992), children's understanding of the physical world still undergo significant changes in childhood (McCloskey, 1983). Previous work has suggested that relevant alternatives may be more easily generated, and have greater affective impact, the closer the outcome is to the goal (e.g., Ong, Goodman, & Zaki, 2015). We can specifically test this hypothesis in our bowling scenario by manipulating the ball's trajectory: imagine an agent who just barely misses getting a strike versus one who misses by a lot. Further, we can consider the possibilities that an agent's desires and prior experiences might also allow children to easily generate relevant alternative outcomes.

Finally, we note several connections between our results and prospect theory (Kahneman & Tversky, 1979). First, adults and children's ratings suggest that their inferences of others' emotions stem from a sensitivity to gains and losses, as opposed to just the final outcomes. Rather than waiting until the final outcome and making inferences based on the outcome, it is possible that people continuously keep track of people's expectations (and presumably their emotions) as the event unfolds, and dynamically adjust their inferences. It would be interesting to design a continuous measure of inference of affective states (e.g., Zaki, Bolger, & Ochsner, 2008) that we can use with children, to study how children make inferences over an extended period of time.

Our findings show an interesting asymmetry between the amount of change for the low expectation and high expectation characters; this might be potentially related to the asymmetric discounting of future losses and future gains. At first glance, prospect theory might seem to suggest the opposite; the gain curve is concave, while the loss curve is convex and steeper. However, this refers to people's predictions about their future losses and gains; for improbable outcomes that were unexpectedly realized, it might have exactly the opposite effect as seen in our data. Recent work has started to apply the concept of a naive utility calculus (Jara-Ettinger, Gweon, Tenenbaum, & Schulz, 2015) to understand how we reason about other minds, and it would be interesting to apply a more general notion of utility calculus in understanding how we reason about others' emotions.

In summary, affective cognition is a critical skill for appropriately and proactively responding to others' emotions, intervening on the world and other people to help them feel better, and critically, avoiding actions that would make others feel worse. These are all incredibly important social capacities, and we see impairments of these capacities in various developmental disorders or psychiatric disorders. We hope that our study will inspire more studies on children's affective reasoning and shed light on helping children develop these skills, better understand their own emotions, and build stronger relationships with others.

Acknowledgments

We thank Andrew McCabe and Willow Bailon for help with data collection and coding, Josh Tenenbaum and Jamil Zaki for helpful discussions, and Bing Nursery School, Palo Alto Junior Museum and Zoo, and the families who participated. This work was supported in part by an A*STAR National Science Scholarship to DCO and by a Varieties of Understanding grant from the John Templeton Foundation to HG.

References

Baillargeon, R., Scott, R. M., & He, Z. (2010). False-belief understanding in infants. *Trends in cognitive sciences*, *14*(3), 110–118.

Barden, R. C., Zelko, F. A., Duncan, S. W., & Masters, J. C. (1980). Children's consensual knowledge about the experiential determinants of emotion. *Journal of Personality and Social Psychology*, *39*(5), 968.

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*(2), 413–426.

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*(3), 569–585.

Gnepp, J., & Hess, D. L. (1986). Children's understanding of verbal and facial display rules. *Developmental Psychology*, *22*(1), 103.

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*(6), 1853–1868.

Hood, B., Cole-Davies, V., & Dias, M. (2003). Looking and search measures of object knowledge in preschool children. *Developmental psychology*, *39*(1), 61.

Jara-Ettinger, J., Gweon, H., Tenenbaum, J. B., & Schulz, L. E. (2015). Children's understanding of the costs and rewards underlying rational action. *Cognition*, *140*, 14–23.

Kahneman, D., & Tversky, A. (1979). Prospect Theory: an Analysis of Decision under Risk. *Econometrica*, *47*(2), 263-292.

Keen, R. (2003). Representation of objects and events why do infants look so smart and toddlers look so dumb? *Current Directions in Psychological Science*, *12*(3), 79–83.

McCloskey, M. (1983). Naive theories of motion. In D. Gentner & A. L. Stevens (Eds.), *Mental models* (pp. 299–324).

Ong, D. C., Goodman, N. D., & Zaki, J. (2015). Near-misses sting even when they are uncontrollable. In *Proceedings of the Thirty-Seventh Annual Meeting of the Cognitive Science Society.* (pp. 1775–1780).

Ong, D. C., Zaki, J., & Goodman, N. D. (2015). Affective cognition: Exploring lay theories of emotion. *Cognition*, *143*, 141–162.

Phillips, A. T., Wellman, H. M., & Spelke, E. S. (2002). Infants' ability to connect gaze and emotional expression to intentional action. *Cognition*, *85*(1), 53–78.

Skerry, A. E., & Spelke, E. S. (2014). Preverbal infants identify emotional reactions that are incongruent with goal outcomes. *Cognition*, *130*(2), 204–216.

Spelke, E. S., Breinlinger, K., Macomber, J., & Jacobson, K. (1992). Origins of knowledge. *Psychological review*, *99*(4), 605.

Weisberg, D. P., & Beck, S. R. (2010). Children's thinking about their own and others regret and relief. *Journal of Experimental Child Psychology*, *106*(2), 184–191.

Weisberg, D. P., & Beck, S. R. (2012). The development of children's regret and relief. *Cognition & emotion*, *26*(5), 820–835.

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*(2), 191–214.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: the truth about false belief. *Child development*, *72*(3), 655–684.

Wu, Y., Baker, C., Tenenbaum, J., & Schulz, L. (2014). Joint inferences of belief and desire from facial expressions. In *Proceedings of the Thirty-Sixth Annual Meeting of the Cognitive Science Society.* (pp. 1796–1801).

Yuill, N. (1984). Young children's coordination of motive and outcome in judgements of satisfaction and morality. *British Journal of Developmental Psychology*, *2*(1), 73–81.

Zaki, J., Bolger, N., & Ochsner, K. (2008). It takes two: the interpersonal nature of empathic accuracy. *Psychological Science*, *19*(4), 399–404.