"If only Santa had one more present": Exploring the development of near-miss counterfactual reasoning

Desmond C. Ong[1,2] (dco@comp.nus.edu.sg), Mika Asaba[3] (masaba@stanford.edu), Hui Yan Lim[4] (limhuiyan@nus.edu.sg), Patricia Chen[4] (patchen@nus.edu.sg), Hyowon Gweon[3] (hyo@stanford.edu)

[1]Department of Information Systems and Analytics, National University of Singapore
[2]Institute of High Performance Computing, A*STAR Singapore
[3]Department of Psychology, Stanford University
[4]Department of Psychology, National University of Singapore

Abstract

Near-misses sting: As adults, we intuitively understand that someone who just missed a desirable outcome (near-miss) feels worse than someone who missed by a far margin (far-miss). What cognitive capacities support these intuitions, and how do they emerge in early childhood? We presented adults (n=42) and six- to eight-year-olds (n=91; pre-registered) with various near-miss scenarios. We found that (1) adults generally infer that a near-miss character would feel worse than a far-miss character, (2) yet their inferences vary depending on the context, and (3) children show a strikingly different pattern from adults, robustly choosing the far-miss character as feeling worse. The tendency to judge the near-miss character as feeling worse increased with age, but even 8-year-olds were still below chance. These patterns raise the possibility that young children start with a distance-based bias that gradually gets replaced by adult-like inferences that involve counterfactual reasoning.

Keywords: Emotion Reasoning; Development; Counterfactual Reasoning

Introduction

Imagine two people, Ben and Ray, who both missed their flights. Ben missed his flight by just a few minutes, and Ray missed his flight by half an hour. Who feels worse? Even though both of them failed to get on their flights, we intuitively understand that Ben might feel worse than Ray; Ben was *almost there* and he *could have* made it if only he had left a few minutes earlier, taken a different route, or run just a little faster. Why do we have this intuition? The current work explores the cognitive capacities that underlie these inferences and how these intuitions emerge in early childhood.

The "sting" of near-misses has been demonstrated in classic studies with adults. For instance, in a scenario similar to the one above, people judged that someone who missed their flight by 5 minutes feels much worse than someone who missed it by 30 minutes (Kahneman & Tversky, 1982); when people are presented with pictures of two Olympic medallists on the podium—one silver and the other bronze—they judge the silver medalist to feel worse even though they won a better prize (Medvec, Madey, & Gilovich, 1995). These studies and many others have consistently found that despite neither agent achieving the desired goal, people judge those who experience near-misses to feel worse than those who experience far-misses, across various types of distance, such as physical distance, temporal proximity, and even the numerical closeness of a lottery ticket to the winning number (e.g., Meyers-Levy & Maheswaran, 1992; Ong, Goodman, & Zaki, 2015).

In order to successfully judge that a near-miss agent feels worse than a far-miss agent, one needs to understand not just *what happened* but also *what could have happened*. Indeed, one theoretical proposal is that individuals who experience near-misses are more likely to generate close counterfactual scenarios in which they successfully achieve their goal; these counterfactual outcomes are compared to the current outcome, leading to a more negative affective reaction (Roese & Olson, 2014; Kahneman & Varey, 1990; Zeelenberg, Van Dijk, Manstead, & der Pligt, 1998). While this proposal concerns first-person experience of near-miss scenarios, similar reasoning can be applied to third-person emotion attribution; we generate an alternative outcome that the character could have experienced, and infer how the character might feel given the contrast between current and alternative outcomes. Thus, prior theories have generally appealed to counterfactual reasoning—the ability to reason about *what could have been*—to explain how we reason about our own and others' emotions in near-miss scenarios (e.g., Ong, Goodman, & Zaki, 2015).

If counterfactual reasoning is critical to understanding emotional responses in near-miss scenarios, one might expect that young children's judgments in these scenarios would parallel the development of counterfactual reasoning. Although prior work has argued that counterfactual reasoning does not develop until the early school years, (e.g., Beck, Robinson, Carroll, & Apperly, 2006), more recent work suggests that such protracted development may be due to the high verbal demand of counterfactual language (e.g., "What would have happened if he had gone the other way?"); studies using simpler scenarios and minimally verbal dependent measures have found earlier success in counterfactual reasoning, as early as preschool-aged children (e.g., Bridgers, Yang, Gerstenberg, & Gweon, 2020; Nyhout & Ganea, 2019).

Critically however, inferring who feels worse in near-miss scenarios involves another step: emotion attribution. In addition to generating the counterfactual outcome that is more positive than the current outcome, children must compare the two and infer that it feels worse to have just missed the goal than to have missed it by far. Despite the early-emerging sensitivity to others' facial expressions (Wu, Muentener, & Schulz, 2017; Walle, Reschke, & Knothe, 2017; Ruba, Meltzoff, & Repacholi, 2019), the ability to reason about how others might feel given their underlying mental states (e.g., desires, beliefs) develops relatively later in preschool years (Asaba, Ong, & Gweon, 2019; Doan, Friedman, & Denison, 2020, in press; Lara, Lagattuta, & Kramer, 2019; Lagattuta, Wellman, & Flavell, 1997; Pons, Harris, & de Rosnay, 2004). Critically, previous studies suggest that counterfactual considerations do not factor into children's judgments of others' counterfactual emotions (regret and relief) until even later in development, at 7 years of age or older (Ferrell, Guttentag, & Gredlein, 2009; Weisberg & Beck, 2010).

If children cannot use counterfactual reasoning to infer emotions, how might they respond to Ben and Ray's example above? One possibility is that children choose either character equally, given that both eventually missed their flights. Another possibility, however, is that children choose the far-miss character as the one who feels worse (e.g., judging that Ray, who missed his flight by 30 minutes, feels worse than Ben who missed it by 5 minutes). Such responses may reflect a relatively simple bias that depends on the closeness of the current outcome to the desired goal (henceforth, "distance bias"); given a positive goal outcome, the closer one gets to the goal, the happier they would be, despite the eventual failure. Prior work with 5-7 year-old children (Guttentag & Ferrell, 2004; McCloy & Strange, 2009) has suggested an alternative account that makes similar predictions to the distance bias heuristic: children rely on a "summation" strategy where agents' emotions in the actual and counterfactual states are "summed", such that an agent who experiences a negative outcome (i.e., chose a box without a prize) but could have achieved a positive outcome (i.e., chose a box with a prize) is judged to feel better than an agent who had to experience a negative outcome regardless (i.e., neither box had a prize). Thus, this work suggests that while children are able to consider alternative, counterfactual states in judging emotions, their ability to appreciate how an alternative positive outcome would make a character feel *worse* seems to be relatively late-developing.

In light of these possibilities, we had two main goals for the current study. First, we sought to investigate the variability in adults' near-miss emotion attribution by using a range of scenarios. In particular, we created scenarios that are similar to ones used in classic adult experiments, as well as ones that focus on physical ability or involve other everyday situations that might be familiar to both adults and children. Importantly however, we did not have an *a priori* categorization of these trials or directed hypotheses about differences between scenarios; rather, the goal was to present children a diverse set of scenarios that are similar in their underlying structure but involve different contexts, in order to identify both consistency and variability in children's responses. Second, we compared children's responses to adult's to examine the emergence of an adult-like understanding of near-miss scenarios. Given that prior work has found varying ages of success in counterfactual reasoning, we also recruited a relatively wide age range—from age 6 to 8—to identify potential developmental change. To see if children's success is accompanied by explicit appeals to counterfactual outcomes, we additionally collected free-response explanations.

In what follows, we describe our experimental procedure with adults and children. We then present the findings with an exploratory discussion about how the near-miss effect may be mediated by situational factors (i.e., the reason for missing the outcome) and speculate on what contextual factors might make counterfactual explanations more accessible.

Methods

Adult Sample. We recruited n=49 American adults online on Amazon Mechanical Turk. We excluded 5 participants who failed an attention check and 2 participants who gave the incorrect answer to the practice trial, leaving a final sample of n=42 participants (M_{Age}(SD) = 37.1(11.1) years; 17 women, 24 men, 1 non-binary).

Child Sample. We recruited n=91[1] six- to -eight year-olds (M_{Age} (SD) = 7.45 years; Range: 6.0-8.9; 30 six year-olds, 31 seven year-olds, and 30 eight year-olds; 47 girls, 44 boys). Data collection was done in a university lab. All participants were native English speakers. Hypotheses, key analyses, and sample size were pre-registered on OSF[2].

Materials

For adults, cartoon images and simple animations were presented via Amazon MTurk. For children, the task was administered using Keynote slides on a computer tablet. There were 8 trials total, each of which involved two characters. The characters were always of the same gender (i.e., two boys or two girls), with an equal number of trials having boy vs. girl characters; character names were displayed on the slide to reduce memory demands. After the experiment was over, participants were presented with a certificate of participation and an age-appropriate toy as a token of appreciation.

Procedure

Practice trial. Children were tested in a private, quiet room in a university lab. In the practice trial, participants were introduced to two characters bowling, who each had a goal of knocking down all of the bowling pins. One bowled a strike and knocked down all the pins, while the other knocked down no pins (similar to the Strike and Gutter trials in Asaba et al., 2019). Participants were asked: "Who feels worse about this bowling game?" All participants answered this practice trial correctly, choosing the agent who knocked down no pins.

Test trials. Children underwent all eight test trials (in one of two orders). In each trial, two agents wanted to achieve the same goal (e.g., get a present from Santa) but ultimately were unsuccessful. Critically, we manipulated how far each agent

[1] We pre-registered n=90 children, but inadvertently collected an additional seven-year-old.

[2] `bit.ly/osf-childnearmiss` ; data: `osf.io/urwds`

Scenario	Characters' Goal	Outcome for near-miss character	Outcome for far-miss character
Soccer	Kick ball into goal	Missed by a little bit	Missed by a lot
Ball Throw	Throw ball into basket	Missed by a little bit	Missed by a lot
Toy Search	Choose 1 of 5 cabinets to find a toy	The toy was in the next cabinet over	The toy was three cabinets away
Card	Randomly receive 3 red cards to win prize	Received 2 red and 1 yellow card	Received 1 red and 2 yellow cards
Cookies	Stand on a chair to reach cookies on a shelf	Cookies were still one shelf higher	Cookies were still two shelves higher
Bus	Walk to catch the bus	Was close to the bus when it left	Was far away from the bus when it left
Santa	Wait in line for a present from Santa	Next in line when presents ran out	Fifth in line when presents ran out
Slide	Wait in line to play on a playground slide	Next in line when break was cancelled	Fifth in line when break was cancelled

Table 1: Scenarios used for children and adults. In each scenario, two characters had the same goal but achieved a different outcome: either very close to the goal (near-miss character) or far from the goal (far-miss character). The notion of far vs. near, however, varied across scenarios (physical distance to the goal, probability of the outcome) as well as the reason for why one was nearer or further from the outcome

was from achieving the goal: one narrowly missed (*Near-Miss*), and the other missed by a larger amount (*Far-Miss*). The two agents were shown on the same screen to minimize memory load and facilitate comparison. Importantly, agents did not influence each other in the trial. For example, in the **Soccer** trial, the two characters were on different halves of the screen, kicking balls towards different goalposts. While two trials—Santa (Fig. 1) and Slide—involved characters waiting in the same line, neither character made active choices and therefore could not have affected the other's outcome.

After the outcome for each character was shown, we asked participants which character feels worse, while intentionally specifying the failed outcome; this was to constrain what aspects of the task children were considering. For instance, in the Santa Trial (Fig. 1), we asked: "Who feels worse about not getting a present?" If children said "both" or "neither", we prompted them again to choose one person who feels worse. Following the test question, children were prompted to provide a verbal explanation for their response: "Why do you think Emma feels worse?"

We designed eight different trials across a variety of contexts. Several were inspired by prior work in the near-miss literature: The **Card** trial was similar to Ong, Goodman, and Zaki (2015), **Toy Search** trial was similar to Doan et al. (in press), and the **Bus** trial was similar to Kahneman and Tversky (1982). We also created trials involving physical skill (**Soccer** trial, **Ball Throw** trial), reaching for an object (**Cookies** trial), and waiting in line (**Santa** trial, **Slide** trial), to broaden the range of contexts presented to children. See Table 1 for descriptions of each trial and the final outcomes for each agent.

Pre-registered hypotheses

We pre-registered three statistical hypotheses for the child sample, predicting that:

1. Six-year-old children would reliably (compared to chance) judge the *Far-Miss* character to feel worse.
2. Older children are more likely to judge the *Near-Miss* character as feeling worse (a linear effect of age on response).
3. By eight-years-old, children would reliably judge the *Near-Miss* character to feel worse. [**not supported*]

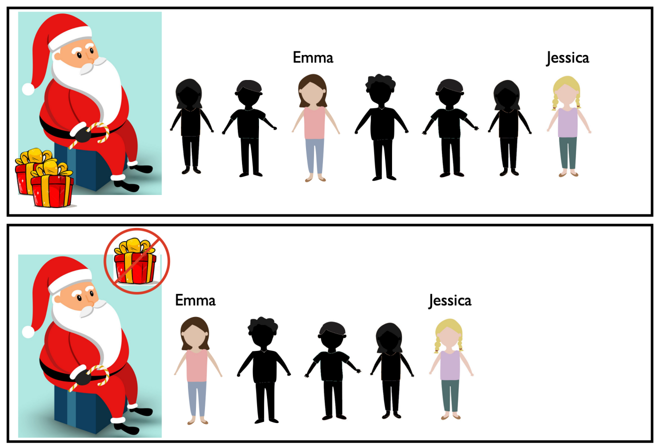

Figure 1: Screenshots from an example trial (Santa). Script: "Emma and Jessica are waiting in line for presents from Santa. Santa has run out of presents! Emma was next in line, while Jessica was fifth in line. Who feels worse about not getting a present?"

Results

Figure 2 shows the results in all 8 scenarios for both adults and children. For child data, the data points represent results averaged across age-groups.

Adults As we expected from prior research with adults on near-miss judgments (Ong, Goodman, & Zaki, 2015; Kahneman & Tversky, 1982), adults in our study generally judged the *Near-Miss* character as feeling worse than the *Far-Miss* character (mean = 74.1%, SD = 24.1%; signed-rank test comparing proportion to 50%, V=643, $p < .001$).

Children Overall, collapsed across all the age-groups, children showed the opposite pattern: They were more likely to judge the *Far-Miss* character as feeling worse ($M_{\text{near-miss}}$ = 11.4%, SD = 31.8%). The results from 6-year-olds confirmed our Pre-registered Hypothesis 1: They chose the *Near-Miss* character as feeling worse 9.6% (SD=16.0%) of the time, which was significantly below chance (planned one-sided signed-rank test comparing against 50%, V=3.5, $p < .001$).

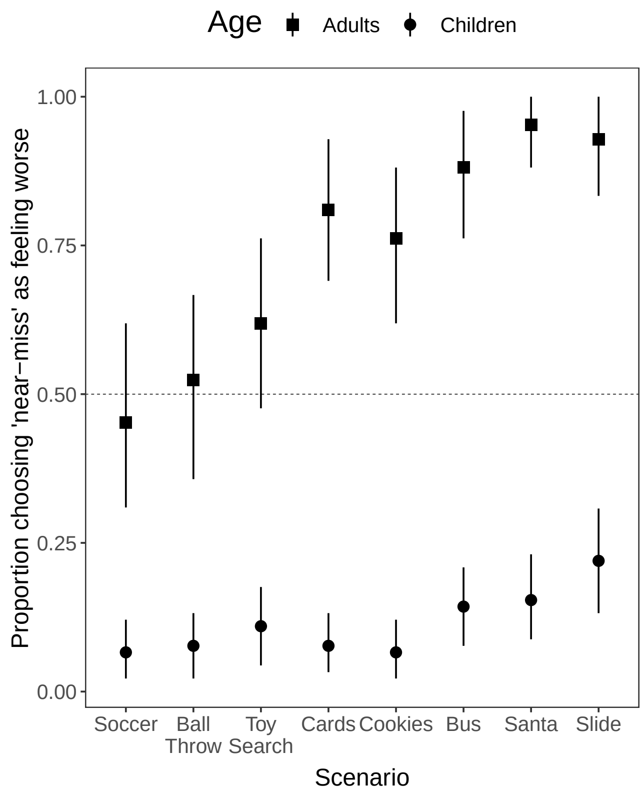

Figure 2: Proportion of children (circle) and adults (square) choosing the near-miss character as feeling worse for each scenario. Error bars are 95% confidence intervals.

We then tested our second hypothesis by regressing the proportion of *Near-Miss* responses against participants' age in months (In R syntax, `lm(proportion_nearmiss ~ age)`). Consistent with our prediction, we observed a significant linear effect of age ($b = .0061$, $t(89) = 2.38$, $p = .019$). That is, older children tended to choose the *Near-Miss* character more than younger children. This effect was modest: according to the model, on average, an additional year of age is associated with a (12*0.0061*100% =) 7.3 percentage points increase in the near miss response. See Fig. 3.

However, our third hypothesis was not supported by the data: Even the oldest participants (8-year-olds) chose the *Near-Miss* character only 19.6% (SD=34.4%) of the time, which is significantly lower than chance level (planned one-sided signed-rank test V=62; p=1).

Exploratory Analyses. Three trends stand out from Figure 2. First, there is a stark difference between adults and children in their choice, with adults choosing the near-miss significantly more (rank-sum test, W = 3582, $p < .001$).

Second, despite this difference and the small number of trials, adults and children's responses followed a similar pattern; this was supported by a near-significant correlation between adult and child responses ($r = .71$, $t(6) = 2.45$, $p = .050$), suggesting that for scenarios where adults' responses tend to favor the *Near-Miss* character, children's responses do, too.

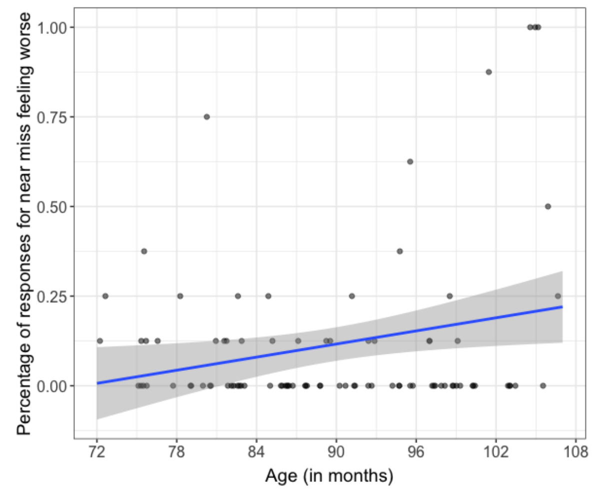

Figure 3: Proportion of trials on which children judged the near-miss agent to feel worse on the vertical axis, as a function of age on the horizontal axis, with best-fit line overlaid. Each data point represents a single child. Data points are semi-transparent with a small horizontal jitter for clarity.

Finally, the trials that involved physical abilities (Soccer and Ball Throw trials) resulted in the weakest choice for the near-miss character. More specifically, the Soccer and Ball Throw scenarios, the left-most two trials in Figure 2, had the lowest proportion of near-miss judgments, across both children (mean across two trials = 7.1%, bootstrapped 95% CI = [3.9%, 11.0%]) and adults (mean = 48.8%, [38.1%, 59.5%]; no difference from chance). We return to this difference in the General Discussion.

Free Response Coding

In order to better understand *why* children made their judgments, we coded their free-response explanations. Two of the authors read through the responses and identified main themes. We identified the first two categories based on pilot work: (i) *Counterfactual*: Explanations that referred to a counterfactual or an alternative state (e.g. "If she had come a little earlier, she would have gotten a present. She feels sad and disappointed"); (ii) *Relative Distance*: Explanations that referred to the closeness, or relative distance (e.g. "She is further than Emma").

In addition to the above two categories, we identified three additional categories from children's responses, for a total of five categories. (iii) *Outcome*: referring to the outcome ("X feels worse because X did not achieve the outcome"), even though both characters experienced the same negative outcome. (iv) *Absolute Position*: A few children ($n_{responses}$=155, or 21.3%) referred to a specific absolute position ("X is last in line" rather than being further back), such as in Fig. 1; this was an unforeseen confound that we plan to remove in follow-up studies. (v) *Others*: Some children gave idiosyncratic ex-

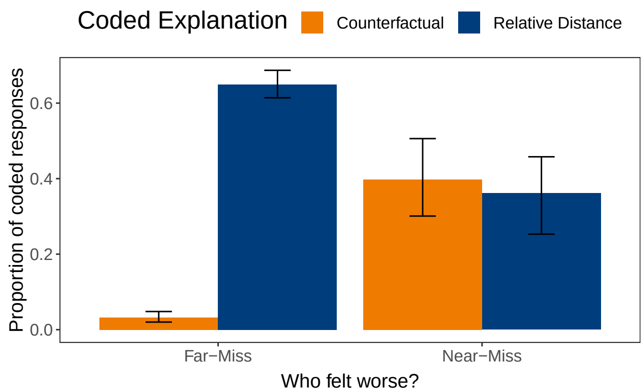

Figure 4: Proportion of coded explanations that referenced a *counterfactual* or *relative distance* by the response that children gave. Error bars are bootstrapped 95% CIs.

Coding Category	Number of Instances (%)
Counterfactual	54 (7.4%)
Relative Distance	449 (61.7%)
Outcome	144 (19.8%)
Absolute Position	155 (21.3%)
Others	38 (5.2%)

Table 2: Results of coding children's free responses. Note that some responses were coded in more than one category.

planations from their personal experience ("I like soccer"), referenced a lack of ability ("X was not good at throwing"), or could not fit into any of the other categories.

Two additional coders independently coded all the responses. Each response could belong to more than one category, if applicable. Coders achieved high percent-agreement rates from 93% to 98%, and Cohen's κ ranging from .79-.94. Coding disagreements were resolved after discussion. The full distribution of coding categories is given in Table 2.

We predicted that children who provided explanations that involve counterfactuals would be more likely to judge the *Near-Miss* character as feeling worse (because the character was a shorter counterfactual away from the outcome); conversely, children who provided explanations that refer to relative distance would be more likely to judge the *Far-Miss* character as feeling worse (because the character was further away). We entered these two coding categories into a mixed-effects logistic model to predict giving a *Near-Miss* response, with random effects by participant and by scenario. In R syntax[3], this was: `glmer(response ~ counterfactual + distance + (1|id) + (1|trial))`.

Indeed, we found that participants who gave a counterfactual response were much more likely to judge the *Near-Miss* character as feeling worse ($b = 3.57$, $z = 4.40$, $p < .001$). By contrast, participants who gave a relative distance response were much more likely to judge the *Far-Miss* character as feeling worse ($b = -1.10$, $z = -2.64$, $p = .008$). In Fig. 4, we can see these two main effects, such that children who gave *Far-Miss* responses overwhelmingly referenced *relative distance* explanations, while children who gave *Near-Miss* responses gave similar proportions of *counterfactual* as *relative distance* explanations.

General Discussion

In this pre-registered study, we examined adults' and children's abilities to provide emotion judgements in "near-miss" scenarios, with a particular focus on using these scenarios to study children's developing ability to incorporate counterfactual judgments into their emotion inferences. Rather than assessing participants' judgments in a single scenario, we collected their judgments across 8 different scenarios; all scenarios had the same underlying structure (two agents, both of whom failed to achieve a positive outcome, but one was closer to achieving the goal (*Near-Miss*) than the other (*Far-Miss*)), but were situated in different contexts.

While adults generally reported that the *Near-Miss* agent felt worse, there was significant variability across scenarios. In contrast, six to eight-year-old children generally reported that the *Far-Miss* agent felt worse; while their judgment became more adult-like with age, even the oldest group (8-year-olds) was reliably below chance. Although even the oldest children in our sample did not judge the near-miss character as feeling worse, we did find evidence of a gradual developmental change between 6 to 8 years of age; older children gave relatively more adult-like responses, which were accompanied by explanations referencing counterfactuals. Indeed, there were several eight-year-olds who gave *Near-Miss* responses on every trial (Fig. 3), suggesting that this developmental shift could occur abruptly around the early schooling years, with substantial individual variability.

Children's failures in these near-miss scenarios stand in a rather stark contrast to recent work showing relatively early success on seemingly similar scenarios. For example, research suggests that by age three or four, children are capable of representing two possible outcomes of an event (Redshaw & Suddendorf, 2016; Leahy & Carey, 2020). Relatedly, five-year-olds thought that a bowler whose ball was initially heading towards the gutter but curved back to hit half of the pins would feel better than another bowler whose ball was initially heading straight down the lane but curved at the last minute to hit only half of the pins (Asaba et al., 2019). Similarly, five-year-olds rated a character that got a yummy gumball from a machine mostly filled with yucky gumballs would be much happier than one who got a yummy gumball from a machine mostly filled with yummy gumballs (Doan et al., 2020). To succeed in these scenarios, children must understand the difference between the expected outcome vs. actual (realized) outcome to understand that even though two people had the same outcome, the one who had a higher expectation to achieve a more desirable outcome may feel worse.

[3] Note that each explanation could be coded in more than one category, so we included `counterfactual` and `distance` as separate (binary) regressors.

Why, then, did children struggle in our near-miss scenarios? One possibility is that the scenarios and stimuli in prior work made it easier for children to represent the alternative (expected) outcome. Alternatively, it is possible that scenarios in these prior studies did not require genuine counterfactual reasoning whereas the current near-miss scenarios did. Consistent with the latter possibility, we found that children at this age often appeal to the relative distance to outcomes, or to counterfacutal outcomes, and that those who mention counterfactual outcomes are more likely to show adult-like responses (i.e., judging the near-miss character to feel worse).

Relatedly, one might also question *why* even the oldest children struggled with this task. Given that 8-year-olds generally do well on counterfactual reasoning tasks (even ones that use counterfactual language), one might have expected to see success in the near-miss scenarios, too. One possibility is that children, despite their competence with counterfactual reasoning, failed to spontaenously generate *relevant* counterfactual outcomes. Another possibility is that children did generate relevant counterfactuals but struggled with drawing inferences about the characters' emotions based on the counterfactual and actual outcomes (Guttentag & Ferrell, 2004; McCloy & Strange, 2009). Some useful insight can be gleaned from the explanation data; children who mentioned counterfactual outcomes were generally more accurate in the near-miss scenarios. Future work could assess counterfactual reasoning and emotion inference separately to better understand the source of difficulty in these tasks.

Another notable aspect of our results is the striking variability in responses across trial types. In particular, the Soccer and Ball Throw trials had the lowest proportion of choices for the *Near-Miss* agent as feeling worse, across children and adults. Why might this be the case? One possibility is that people may be making dispositional attributions about the agents' underlying abilities or competence; that is, children and adults may have reported that the *Far-Miss* character feels worse because their poor performance indicates low competence. Note however that *Near-Miss* effects have been found in scenarios that are ability-based (i.e., Olympic medalists, see Medvec et al., 1995), suggesting that ability itself is not the key mediator of this effect; various contextual factors can invite participants to generate different alternative explanations about the reason for the characters' failure, including judgments of competence or the probability of achieving the goal. Although the current work was not designed to identify exactly what factors are at work, the variability across scenarios provide us initial insights into these factors.

So far we have discussed counterfactual reasoning as a likely prerequisite for our task. One might wonder, however, whether counterfactual reasoning is the only route to making the "correct" emotion attribution in this task. An alternative possibility is that adults are reasoning about the character's *expectations* about potential outcomes (i.e., what will happen), rather than counterfactual outcomes, and compare these expectations to the actual outcome. For instance, in the Santa trial, the agent closer to Santa may be perceived to have higher expectations of getting a present compared to the agent further away from Santa. After both agents fail to achieve their outcome, the agent with higher expectation would be judged to feel worse. Note that this possibility may be insufficient to fully explain the stark differences between adults and children we found in the current study; by 5 years of age, children's expectation-based reasoning about others' emotions start to look more adult-like (Asaba et al., 2019), and continues to develop between age 5 - 8 (Lara et al., 2019). The fact that children rarely mentioned expectations in their explanations also makes this possibility somewhat less likely. However, the current study is not ideally suited to test hypotheses about underlying strategies (counterfactual-based vs. expectation-based); most of these scenarios can be answered using either strategy, and some scenarios also involve expectations that change rapidly during the course of the trial. Future work can isolate these processes (reasoning about agents' expectations vs. counterfactual outcomes) by matching the characters' expectations throughout the stories.

In sum, these results suggest that although near-misses sting, the "sting" may indicate a non-trivial developmental feat: Understanding what could have happened, and being able to reason about how others might feel. Although school-aged children look almost adult-like in many social cognition tasks, these near-miss scenario represent cases where children make robust inferences that are systematically and strikingly different from that of adults. Such systematic developmental differences have inspired decades of work on standard false-belief reasoning; we look forward to future work on the development of counterfactual reasoning and affective cognition (Ong, Zaki, & Goodman, 2015).

Acknowledgments

This work was supported in part by a Singapore Ministry of Education Academic Research Fund Tier 1 grant to DCO.

References

Asaba, M., Ong, D. C., & Gweon, H. (2019). Integrating expectations and outcomes: Preschoolers' developing ability to reason about others' emotions. *Developmental Psychology*, *55*(8), 1680.

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.

Bridgers, S., Yang, C., Gerstenberg, T., & Gweon, H. (2020). Whom will Granny thank? Thinking about what could have been informs children's inferences about relative helpfulness. In *Proceedings of the 42nd Annual Meeting of the Cognitive Science Society.*

Doan, T., Friedman, O., & Denison, S. (2020). Young children use probability to infer happiness and the quality of outcomes. *Psychological Science*, *31*(2), 149–159.

Doan, T., Friedman, O., & Denison, S. (in press). Oh...so close! Children's close counterfactual reasoning and emotion inferences. *Developmental Science*.

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.

Guttentag, R., & Ferrell, J. (2004). Reality compared with its alternatives: age differences in judgments of regret and relief. *Developmental Psychology*, *40*(5), 764.

Kahneman, D., & Tversky, A. (1982). The simulation heuristic. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 201–208). Cambridge University Press.

Kahneman, D., & Varey, C. A. (1990). Propensities and counterfactuals: The loser that almost won. *Journal of Personality and Social Psychology*, *59*(6), 1101.

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*, 1081–1104.

Lara, K. H., Lagattuta, K. H., & Kramer, H. J. (2019). Is there a downside to anticipating the upside? children's and adults' reasoning about how prior expectations shape future emotions. *Child Development*, *90*(4), 1170–1184.

Leahy, B. P., & Carey, S. E. (2020). The acquisition of modal concepts. *Trends in Cognitive Sciences*, *24*(1), 65–78.

McCloy, R., & Strange, P. (2009). Children's understanding of counterfactual alternatives. In *Proceedings of the 31st Annual Meeting of the Cognitive Science Society* (pp. 1627–1632).

Medvec, V. H., Madey, S. F., & Gilovich, T. (1995). When less is more: Counterfactual thinking and satisfaction among Olympic medalists. *Journal of Personality and Social Psychology*, *69*(4), 603.

Meyers-Levy, J., & Maheswaran, D. (1992). When timing matters: The influence of temporal distance on consumers' affective and persuasive responses. *Journal of Consumer Research*, *19*(3), 424–433.

Nyhout, A., & Ganea, P. A. (2019). Mature counterfactual reasoning in 4-and 5-year-olds. *Cognition*, *183*, 57–66.

Ong, D. C., Goodman, N. D., & Zaki, J. (2015). Near-misses sting even when they are uncontrollable. In *Proceedings of the 37th 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.

Pons, F., Harris, P. L., & de Rosnay, M. (2004). Emotion comprehension between 3 and 11 years: Developmental periods and hierarchical organization. *European Journal of Developmental Psychology*, *1*(2), 127–152.

Redshaw, J., & Suddendorf, T. (2016). Children's and apes' preparatory responses to two mutually exclusive possibilities. *Current Biology*, *26*(13), 1758–1762.

Roese, N. J., & Olson, J. M. (2014). *What might have been: The social psychology of counterfactual thinking*. Psychology Press.

Ruba, A. L., Meltzoff, A. N., & Repacholi, B. M. (2019). How do you feel? preverbal infants match negative emotions to events. *Developmental Psychology*, *55*(6), 1138.

Walle, E. A., Reschke, P. J., & Knothe, J. M. (2017). Social referencing: Defining and delineating a basic process of emotion. *Emotion Review*, *9*(3), 245–252.

Weisberg, D. P., & Beck, S. R. (2010). Children's thinking about their own and others' regret and relief. *Journal of Experimental Child Psychology*, *106*(2-3), 184–191.

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*, *114*(45), 11896–11901.

Zeelenberg, M., Van Dijk, W. W., Manstead, A. S., & der Pligt, J. (1998). The experience of regret and disappointment. *Cognition & Emotion*, *12*(2), 221–230.