Whom will Granny thank?

Thinking about what could have been informs children's inferences about relative helpfulness

Sophie Bridgers, Chuyi Yang, Tobias Gerstenberg, Hyowon Gweon
{sbridge, cayang, gerstenberg, hyo}@stanford.edu
Department of Psychology, Stanford University
450 Jane Stanford Way, Stanford, CA, USA, 94305

Abstract

To evaluate others' actions, we consider action outcomes (e.g., positive or negative) and the actors' underlying intentions (e.g., intentional or accidental). However, we often encounter situations where neither actual outcomes nor intentions provide useful evidence for evaluation but representations of unrealized (counterfactual) outcomes matter. Here we ask whether preschool-aged children consider counterfactual outcomes to evaluate whose action was more helpful. When two agents each caught one of two falling apples (one caught it above a trash can and the other above a fruit basket), children chose the former as the one who should be thanked (because otherwise the apple would've fallen into the trash). When the agents caught crushed cans, however, children made the opposite choice, choosing the agent who caught the can over the fruit basket. Even though preschoolers typically struggle with counterfactuals, children in our task readily engaged in such reasoning in the context of social evaluation.

Keywords: causal reasoning; social cognition; Theory of Mind; counterfactual simulation; prosocial actions.

Introduction

Humans are social creatures. Much of our daily thought is dedicated to what others do and why they do what they do. Did the stranger open the door for me, or was he opening it for himself and I just happened to approach at the same time? My child is trying to help me clean the kitchen, though in reality she is making more of a mess. We were supposed to work on a conference submission together, but some authors worked harder than others. These social evaluations not only inform our judgments about who deserves credit or blame, but also guide our interactions with others (Rai & Fiske, 2011).

As adults, we evaluate others' actions based not only on the outcomes they bring about but also on their underlying intentions (Cushman, 2008; Schächtele, Gerstenberg, & Lagnado, 2011; Young, Cushman, Hauser, & Saxe, 2007). While the ability to distinguish helpful versus harmful agents emerges early in life (Hamlin, Wynn, & Bloom, 2007), research suggests that the ability to consider intentions (and not just outcomes) when making moral evaluations develops across early childhood (e.g. Cushman, Sheketoff, Wharton, & Carey, 2013; Nelson, 1980). Given that intentions are abstract and latent, they ought to be harder, in principle, to consider than clearly observable outcomes. In general, the ability to consider intentions in moral judgments has been considered a window into the development of mental-state reasoning (Baird & Moses, 2004; Cushman et al., 2013; Yuill & Perner, 1988). Thus, prior literature on moral reasoning has typically focused on children's capacity to consider intentions when such information is provided explicitly in context.

Critically however, in many real-world scenarios, we don't get direct, explicit information about the outcomes of others' actions nor their intentions. Actions do not always yield a clearly positive or negative state (e.g., preventing a door from closing), nor do people always broadcast the intentions behind such actions. If we were to consider only observable, explicitly available information, the moral status of many actions we encounter would likely remain ambiguous. Yet, as adults, our evaluations go beyond considerations of actual outcomes or actors' intentions; given an action and its outcome, we also think about the *potential, unrealized outcomes* of inaction (i.e., what could have happened if the actor had not intervened) to evaluate the praiseworthiness or helpfulness of the action.

Consider the scenario in Fig. 1. Granny accidentally dropped two apples on the table, and each apple rolled off the table to either side. One apple was headed for a trash can and the other for a fruit basket. However, just before that happened, Susie and Annie caught the apples! Susie caught

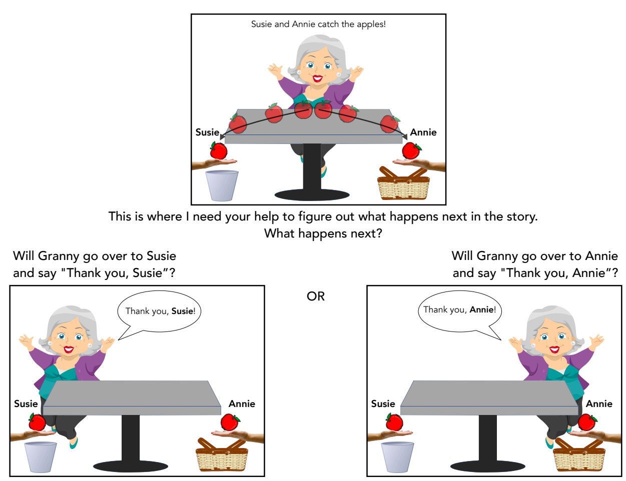

Figure 1: **Experiment 1 Apples Condition**. Top image: Susie and Annie catching the apples (displayed on the laptop screen). Bottom images: laminated cards to which children can point to indicate whom Granny will thank. Trajectories of rolling apples are displayed here for reference but were not present on this slide for children.

the apple that would have fallen into the trash can, and Annie caught the apple that would have fallen into the fruit basket. Who was more helpful? Susie, or Annie?

In this scenario, the two actors performed identical actions, presumably with the same intention of catching the apples. Nonetheless, you might have the intuition that Susie's action was more helpful since she prevented the apple from falling into the trash can. Annie's action wasn't as helpful, because it wouldn't have been so bad if the apple had fallen into the fruit basket. In order to evaluate Susie's action as more praiseworthy than Annie's, it is insufficient to consider their actions, immediate intentions, or the observed outcomes. Instead, we must reason about what would have happened if either of them had not acted. While intuitive, this inference is actually quite complex. The observer must go beyond the observed outcome and consider alternative possibilities that could have otherwise occurred (i.e., counterfactual thinking). Counterfactual thinking involves a comparison between what actually happened, and what would have happened if the action had not taken place—that is, a comparison between reality and a simulated possibility based on alternative past events (Byrne, 2016; Lewis, 1973).

Prior work suggests that children can reason about the expected (but not yet realized) outcomes of actions and make prosocial decisions accordingly. Three-year-olds override a person's direct request for a tool when it is inappropriate for the task and provide the appropriate tool (Martin & Olson, 2013). Five- to 7-year-olds consider a learner's expected utility of being taught vs. discovering information on one's own to preferentially teach rewarding, difficult-to-discover information (Bridgers, Jara-Ettinger, & Gweon, 2020). Consistent with these findings, recent work on development of modal representations suggests children's ability to simultaneously represent two simulated possibilities appears to develop from age 3 to 5 (Leahy & Carey, 2020).

However, inferring whom Granny would thank involves a retrospective evaluation of two preventative actions. Rather than comparing two future outcomes (i.e., future hypothetical reasoning), one must compare an observed outcome of an action that has already occurred (i.e., the current state of the world) against a counterfactual outcome that could have happened in the absence of the action. The evidence for counterfactual reasoning during the preschool years is mixed. Some research suggests that children cannot engage in counterfactual reasoning until middle to late childhood (Beck, Robinson, Carroll, & Apperly, 2006; Rafetseder, Schwitalla, & Perner, 2013), while other research claims to have found success at age 4 (Guajardo & Turley-Ames, 2004; Harris, German, & Mills, 1996; Nyhout & Ganea, 2019). Given these findings, it is unclear whether children would be able to determine which of two agents was more helpful based on what might have happened if they hadn't helped.

Critically however, the complexity of the sentence *What might have happened if they hadn't helped?* reveals a key challenge in studying the development of counterfactual reasoning. Much existing work examining children's counterfactual reasoning involves the use of explicit verbal prompts with a particular grammatical tense: past subjunctive (Kuczaj & Daly, 1979). Unsurprisingly, counterfactual success correlates with language ability and mastery of the past subjunctive (Harris, 1997). The linguistic demand of these verbal tasks raises the possibility that children can spontaneously engage in such reasoning, especially when they need to use it to solve a problem (such as a social evaluation) that does not involve answering a question in the past subjunctive tense.

In the Granny scenario (Fig. 1), answering who was more helpful requires reasoning about what would have occurred if the friends had not intervened. Thus, we may be able to capture early counterfactual reasoning abilities without using grammatically complex verbal prompts.

Here, we ask whether 3- to 5-year-old children are able to make relative judgments of others' helpfulness based on representations of unrealized outcomes. Children were presented with the story of Granny who drops objects (either apples or empty, crushed cans) onto a table that roll off the table and appear as if they are about to fall into either a fruit basket or a trash can, but Granny's friends, Annie and Susie, catch the objects before they fall into these receptacles. Children are then asked whom Granny will thank.

In order to determine whom was more helpful (i.e., who will receive thanks), children have to reason about what would have occurred if the friends had not intervened. Answering this question relies on counterfactual thinking but the question itself does not tax children's linguistic abilities. Furthermore, rather than explicitly signaling the need to reason counterfactually, this design capitalizes on children's intuitive understanding of physics to induce counterfactual thinking (i.e., the simulation they need to carry out is straightforward given their understanding of the physical world; Gerstenberg, Peterson, Goodman, Lagnado, & Tenenbaum, 2017; Kominsky et al., 2019) .

Since children between 3 to 5 years of age exhibit sophisticated causal reasoning (e.g., Gopnik, Sobel, Schulz, & Glymour, 2001; Kushnir & Gopnik, 2005), sensitivity to the helpfulness of different possible actions (Bridgers et al., 2020; Martin & Olson, 2013), and mixed success with counterfactual reasoning (Beck et al., 2006; Nyhout & Ganea, 2019).

Experiment 1

In this experiment, we operationalize helpfulness as who is more likely to receive thanks for their action. This indirect measure (i.e., rather than asking who was more helpful) has several advantages: (1) "Who will Granny thank?" is action-based–arguably simpler and clearer–than helpfulness, which might change in its meaning over early childhood; (2) Anecdotal experience from our prior work indicates that children are hesitant to judge relative helpfulness of agents especially when the agents are visually present and neither agent did something clearly wrong; (3) The thanking event is not necessarily an event children could simulate or predict ahead of

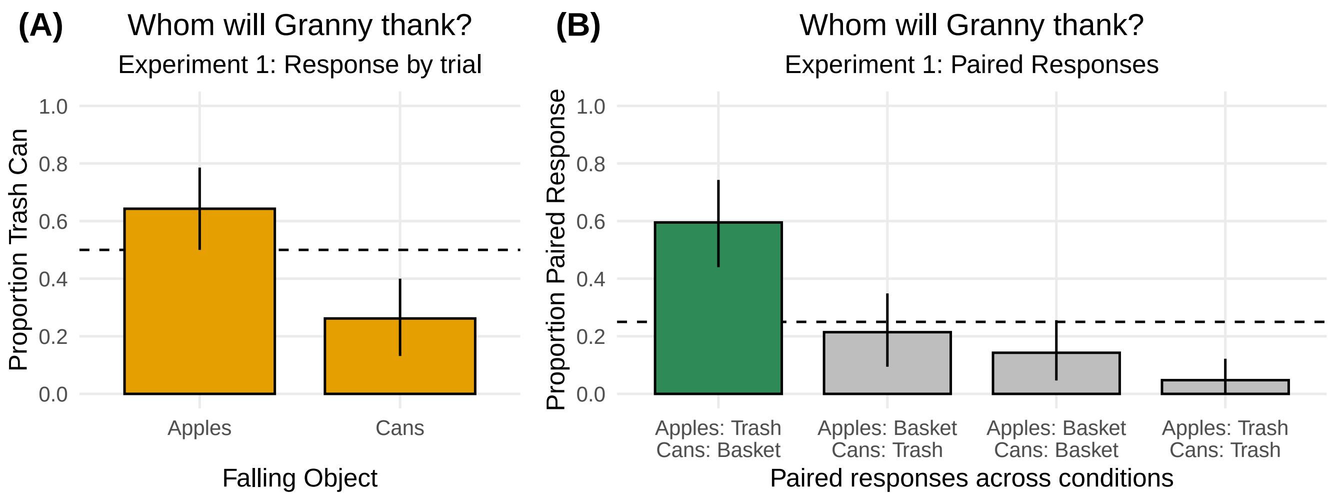

Figure 2: **Experiment 1 Results**. (A) Proportion of children who predicted Granny would thank the agent who caught the object that would have fallen into the trash can. (B) Proportion of each possible paired response across conditions; green bar represents children who answered correctly in both conditions. Error bars are bootstrapped 95% confidence intervals.

time (at least on the first trial), making it more likely that they would need to engage in counterfactual reasoning, rather than future hypothetical reasoning, to appropriately respond.

We predicted that in the apples case, children would predict that Granny would thank Susie, who caught the apple over the trash can (because otherwise it would have fallen into the trash), but in the crushed cans case, that she would thank Annie, who caught the can over the fruit basket (because otherwise it would have contaminated the fruit basket). If children cannot reason about what would have happened if Susie and Annie had not intervened, then they might predict that Granny would be equally likely to thank either agent (regardless of the object's identity) since they both caught an object. Alternatively, if children reason forward from the event of catching the objects or simply associate items with their appropriate receptacles, they might make the opposite predictions as those based on counterfactual reasoning. In the apples case, they might predict that Granny would thank Annie because she is holding the apple over the fruit basket and in the cans case, predict that Granny would thank Susie because she is holding the crushed can over the trash can.

Methods

Participants Forty-two 3-to-5-year-olds ($\text{Mean}_{\text{age}}(SD) = 57.53(6.97)$ months; range: 41.26 - 70.29 months; 52% female) were recruited from Stanford University's Bing Nursery School. The design was within-subject; all children participated in both conditions: Apples condition and Cans condition (order counterbalanced across participants). An additional 7 children were recruited but excluded from analysis due to failing to provide a response (n = 6) and ending the experiment early (n = 1).

Materials A keynote presentation displayed on a laptop was used to narrate three different scenarios of Granny dropping sets of objects onto her kitchen table (2 blue-colored balls, 2 red apples, 2 empty, crushed red cans); the trash can was always on the left and the fruit basked on the right. The helpers, "Susie" and "Annie", were depicted as human hands holding an object over a receptacle with their names displayed in text above. Four 8.5in. x 11in. laminated cards showing Granny thanking either Susie or Annie were also used.

Procedure The study was conducted in a room separate from the children's classrooms. Children sat at a table directly in front of the laptop. Children were told they were going to listen to a story, which the experimenter proceeded to read aloud from the keynote presentation. Children were first introduced to Granny, her kitchen table (which she stood behind), her trash can, and her fruit basket.

World building: Training Scenario. The first scenario was always Granny dropping the two balls. Once dropped, the balls began to simultaneously roll off the table, and the experimenter read: "Oh no! The balls are rolling off the table!". The ball on the left side of the table rolled off the edge and disappeared into the trash can. At the same time, the ball on the right side rolled off and disappeared into the fruit basket. The experimenter asked children to point to where the balls were now. Once children responded, the trash can and fruit basket turned transparent revealing a ball inside each one. If children only pointed to one receptacle, the experimenter asked "Where did the other ball go?". If children did not point to either receptacle, the experimenter revealed the balls and noted that one had fallen into the trash can and the other into the fruit basket. This scenario was designed to help children understand that in the world of Granny's kitchen, objects roll off the table and fall into the receptacles below as opposed to onto the ground or elsewhere (i.e., it clarified the physics of the situation). Regardless of how children responded, they were included in the final analysis.

Evaluating the helpfulness of actions: Test Scenarios. Next, children observed Granny drop two new objects onto her kitchen table, either apples (Apples condition) or cans (Cans condition). Once dropped, these objects too proceeded to roll off the table in tandem, but before they could fall into the receptacles, Susie and Annie appeared and caught the objects (Susie's hand always appeared on the left, catching the object over the trash can, and Annie's hand always appeared on the right, catching the object over the fruit basket). The experimenter narrated: "Susie and Annie catch the apples/cans", and explained to children, "This is where I need your help to figure out what happens next in the story." The experimenter took out the two laminated cards corresponding to the condition and asked, "Will Granny go over to Susie and say 'Thank you, Susie!', or will Granny go over to Annie and say 'Thank you, Annie!'? What happens next?" (whether the Susie or Annie card was described first was counterbalanced in each condition across participants). (Fig. 1) After children responded, the experimenter moved onto the next condition.

The card children pointed to was recorded as their choice. If children did not point, but said a name aloud, then the name stated was recorded as their choice. If children did not respond, even after additional prompting, the experimenter continued with the story. Only children who responded in both conditions, however, were included in analysis.

Results and Discussion

We fit a mixed effects logistic regression predicting children's choice of agent-receptacle pair ("Susie/trash" v. "Annie/basket") with fixed effects of condition ("Apples" v. "Cans"), age (continuous), and their interaction, as well as a random intercept for each subject. This analysis revealed that children were significantly more likely to predict that Granny would thank Susie (who caught an object over the trash can) when that object was an apple than when it was an empty, crushed can (condition: $\beta = -1.625$, $z = -3.410$, $p < .001$, see Fig. 2A). There was no main effect of age nor a condition by age interaction (age: $\beta = -0.133$, $z = -0.248$, $p = .804$; condition*age: $\beta = 0.136$, $z = 0.172$, $p = .863$).

Our main prediction was the condition difference, but we were also curious whether within each condition children's choices would differ from chance; specifically, in the Apples condition would children be more likely to predict that Granny would thank Susie who caught the apple above the trash can than Annie who caught the apple above the fruit basket and vice versa in the Cans condition. We conducted a one-tailed Wald test on the intercept of the logistic regression described above and on the intercept of the same regression re-fit with the Cans condition dummy coded as the reference level. Children indeed were significantly more likely to predict that Granny would thank Susie than Annie in the Apples condition (64.29% predicted Susie/trash; $\beta = 0.589$, $z = 1.826$, $p = .034$) but were significantly more likely to predict Granny would thank Annie than Susie in the Cans condition (73.81% predicted Annie/basket; $\beta = -1.036$, $z = -2.952$, $p = 0.002$). Moreover, children made the correct prediction in both conditions at a level significantly greater than chance (59.52%; two-tailed Binomial test with 25% chance, $p < .0001$, see Fig. 2B).

The only difference across conditions was the identity of the objects that were falling. This difference, however, changed which agent performed a more helpful act by preventing a worse (unrealized) outcome. Children's predictions of whom Granny would thank varied by object and appropriately with the severity of the counterfactual outcome, suggesting that children were sensitive to what would have happened if the agent had not helped. Encouraged by these initial results, we decided to preregister a larger-scale replication to verify their reliability.

Experiment 2: Ongoing Preregistered Experiment

The sample size, inclusion/exclusion criteria, procedure, hypotheses, and analyses for Experiment 2 are preregistered on the Open Science Framework (`https://osf.io/96m2b`). Our planned sample size is 96 participants (approximately 32 3-year-olds, 32 4-year-olds, and 32 5-year-olds). Here, we present the preliminary findings.

Methods

Participants Thus far we have recruited 42 children (n = 15 3-year-olds, n = 23 4-year-olds, n = 4 5-year-olds; $\text{Mean}_{\text{age}}(SD) = 51.59(6.54)$ months, range: 38.83 - 62.01 months, 40% female) from Bing Nursery School. An additional four participants were recruited but excluded from analysis due to parental interference (n = 1), ending the experiment early (n = 2), and insisting Granny would thank both Susie and Annie (n = 1).

Materials We added two new training scenarios. Laminated cards depicting two possible actions Granny could take within each scenario are also used (i.e., four 8.5in. x 11in. laminated cards showing Granny watering either the healthy plant or the thirsty plant, and Granny giving a band-aid to either Max or John as described below). The rest of the slides and the four laminated cards for the test scenarios are the same as in Experiment 1.

Procedure The procedure is identical to that of Experiment 1 with the following modifications.

New Training Scenarios. First, in order to make the task more clear, especially for 3-year-olds, we included two training trials (order counterbalanced across participants) to give children practice predicting what would happen next in the story and selecting a card depicting that prediction. In one training trial, Granny is in her garden where she has a watering can, a "happy, healthy plant" (on the left), and a "sad, thirsty plant" (on the right). Granny only has enough water in her watering can to water one of the plants. Children are asked to predict which plant Granny will water and to indicate their choice by pointing to a picture of Granny watering the healthy plant or a picture of Granny watering the thirsty plant.

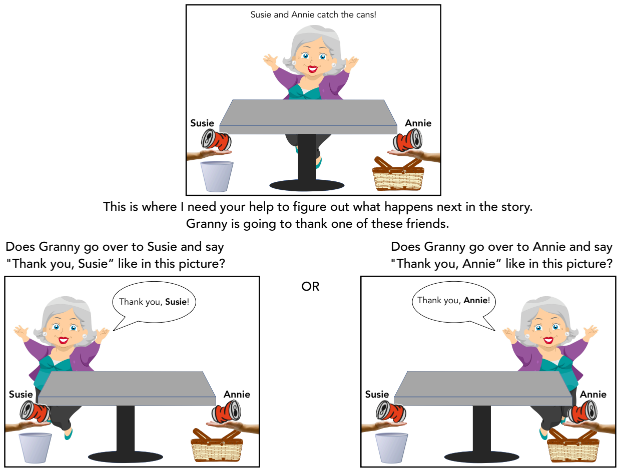

Figure 3: **Experiment 2: Cans Condition**. Top image: Susie and Annie catching the cans (displayed on laptop). Bottom images: laminated cards to which children can point to indicate whom Granny will thank.

In the other training trial, Granny is in her living room where her two grandsons, Max (on the left) and John (on the right), are playing. Granny has a single band-aid and she observes Max fall and hurt his knee. Children are asked to predict to whom Granny will give the band-aid by pointing to a picture of Granny standing next to Max holding the band-aid or a picture of Granny standing next to John holding the band-aid.

The training trials were designed to involve objects and agents, like the test trials, but to be unrelated in content to the test trials. Additionally, the training trials had salient "correct" responses to guide children's responses and build their confidence in making predictions about what would happen next in the story. We hoped this practice would encourage them to respond in the more ambiguous test trials and to more clearly indicate their choice by pointing to a card depicting the next scene (since some children in Experiment 1 verbally stated whom Granny would thank while also pointing at a card showing Granny thanking the opposite person). The rest of the story proceeds in the same way as Experiment 1 (balls training trial followed by Apples and Cans test trials).

Modifications to Test Trials. The second modification from Experiment 1 is how children are prompted to make their prediction about whom Granny will thank. We aimed to clarify that Granny would only thank one friend (since some children in Experiment 1 said she would thank both Susie and Annie) and that children should indicate their prediction by pointing to the entire card rather than a specific character. Here, after Susie and Annie are shown catching either the apples or the cans, the experimenter says: "Okay, this is where I need your help to figure out what happens next in the story. Granny is going to thank one of these friends. Does Granny go over to Susie and say 'Thank you, Susie' like in this picture or does Granny go over to Annie and say, 'Thank you, Annie' like in this picture?" (see Fig. 3).

The third modification is that children are always encouraged to point. The picture children point to is recorded as their choice. If children just say a name aloud but do not point or if they point to the picture depicting Granny thanking the opposite character from whom they named, they are excluded from analysis since the intended response is unclear.

Results and Discussion

Overall, children are succeeding on the new training trials: 90% of children answered correctly on each trial, and this training appears to support children's responding on the test trials: all children have provided a clear response.

The trends observed in the data thus far appear consistent with the results from Experiment 1. Currently, 59.52% of children predict that Granny will thank Susie (who catches objects over the trash) when the object she is catching is an apple, while 47.6% of children predict Granny will thank Susie when the object she is catching is an apple. Moreover, 42.86% of children make the correct prediction in both conditions, which is significantly different from chance (two-tailed Binomial test with 25% chance, $p = .011$). The current effect size appears to be smaller than Experiment 1, but nonetheless promising given that we have fewer participants than Experiment 1 and only have collected 44% of our planned sample.

General Discussion

In this study, we asked children to evaluate the helpfulness of others' actions. However, rather than providing explicit information about the valence of outcomes or actors' intentions, we provided a scenario that could, in principle, elicit representations of prevented (and thus unrealized) outcomes. Successfully answering this question thus required reasoning about what would have happened if an agent had not helped.

Counterfactual reasoning about what would have happened is challenging for young children. Here, however, we find preliminary evidence that 3- to 5-year-olds can determine which one of two actors was more helpful based on the relative negativity of the alternative outcomes that they prevented. In addition to identifying which actions will be more helpful based on the expected consequences for the recipient (Bridgers et al., 2020; Martin & Olson, 2013), these results suggest that children can evaluate the helpfulness of actions that already occurred based on their counterfactual alternatives.

Note that these inferences are quite subtle. Both actors helped by catching identical objects that have just fallen off a table, but prevented different outcomes. The relative severity of the two counterfactual outcomes do differ; an apple falling into a trash can is indeed worse than an apple falling into a fruit basket. However, even this worse outcome is by no means a catastrophe – Granny could just wash off the apple. Likewise, an empty, crushed can falling into a fruit basket is worse than it falling into a trash can, but it is not the end of the world. Despite the subtlety of these scenarios, remarkably, children were able to infer that Granny would be more likely to thank the character who prevented the (slightly) worse outcome.

In many tasks assessing early counterfactual reasoning, the question itself draws children's attention to the past event they should un-do (e.g., "What would have happened if this block was not on the machine?"). However, we did not use such questions (i.e., "What would have happened if Susie had not caught the apple?"). Thus, to reason counterfactually in our task, children must spontaneously identify what event should be undone or altered to initiate the counterfactual simulation (i.e., the actors' catching of the apples, rather than Granny's dropping of the apples).

One might be left unconvinced, however, whether success in our task reflects genuine counterfactual reasoning (i.e., comparing the helpers' actions to what would have happened if they had not acted). Note that the visual scenes–one hand holding an object over the trash can and the other holding the same object over the fruit basket–do not provide any cues to the correct answer. In fact, if children simply simulated forward from that point in the story (instead of engaging in a counterfactual simulation), they might instead make the opposite predictions; they might predict Granny would thank the person holding the apple over the fruit basket and the person holding the can over the trash can, anticipating that these people are about to place the object in the receptacle directly beneath their hands. Likewise, if children were simply associating objects with their corresponding receptacles, they would also make the opposite choice than what is consistent with counterfactual reasoning, because apples go in fruit baskets and empty cans go in trash cans. Finally, children could not answer by simply recalling prior events, because nowhere in the the course of the story do children see apples or cans falling into the trash can or the fruit basket.

However, we note one plausible alternative explanation. During the practice trials, children observed balls roll off the edge of the table into the different receptacles. Thus, even though children never saw the apples or cans actually roll off the table and fall into the receptacles during the test trials, they might have simulated forward from the moment the objects began rolling (rather than initiating a counterfactual simulation when the objects were caught or when asked who Granny will thank), especially on the second test trial. The valence of the future hypothetical outcomes of these simulations might have informed children's inferences about whom Granny will thank.

These simulated outcomes, like the counterfactual ones, have different valences, so the agent who prevents the more negatively valenced outcome is the one whom Granny will thank. Such future hypothetical reasoning implies sophisticated simulations and the ability to compare two possible outcomes, but it may fall short of genuine counterfactual reasoning that involves going back in time, un-doing an event, and reasoning about what could have been otherwise.

For this reason, we are cautious to interpret children's success on this task as evidence for genuine counterfactual reasoning (versus hypothetical thinking). In fact, controlling for the possibility of future hypothetical reasoning is a design challenge for the majority of work examining counterfactual reasoning and most developmental studies showing early competence are not free from similar limitations (see Beck, 2015; McCormack, Ho, Gribben, O'Connor, & Hoerl, 2018; Nyhout & Ganea, 2019; Nyhout, Henke, & Ganea, 2017; Rafetseder & Perner, 2018; Weisberg & Gopnik, 2013).

While preliminary, the current results also speak to the promise of embedding complex reasoning in social situations where children need to appeal to such reasoning to solve a problem. We ask children an indirect yet simple question about whom Granny will thank. This makes it less likely that children's failures are due to verbal demands of the task. Note also that we chose to ask about "thanking" rather than directly probing helpfulness; although one might wonder whether results would have been different if we asked "who is more helpful", we have no a priori reason to suspect this, and asking about thanking offered clear methodological advantages.

Here, we focused on how reasoning about the unrealized, negative outcomes of inaction informs evaluations of the helpfulness of an action. Relatedly, reasoning about prevented (thus unrealized) positive outcomes could inform evaluations of the harmfulness of an action. There is evidence that children's own experience of regret vs. relief follow different developmental trajectories, and that they struggle to attribute these emotions to others' until age 7 (Weisberg & Beck, 2010). Exploring this broader space will provide additional insight on the role of unrealized outcomes in children's social evaluations.

The causal structure of social interactions is often uncertain and complex. Children's ability to evaluate the utility of different actions in the real-world likely will depend on their ability to represent the full causal structure of these actions, as well as children's capacities to consider not just observed outcomes but multiple unrealized outcomes. Indeed, even inferring another person's intention to bring about an outcome can depend on representing potential alternative events (Kleiman-Weiner, Gerstenberg, Levine, & Tenenbaum, 2015). We present initial evidence that children can discern the relative helpfulness of actions not just from what actually happened, but also by thinking about the unfulfilled outcomes the actions prevented. As much as these results inform our understanding of how children evaluate others' actions, our experimental paradigm also offers a promising approach to examining the early development of causal and counterfactual reasoning in social contexts.

Acknowledgments

We thank the staff, families, and children of Bing Nursery School. This work was supported by an NSF GRFP to SB and a McDonnell Scholars Award and a Jacobs Foundation Scholars Award to HG.

References

Baird, J. A., & Moses, L. J. (2004). The role of mental state understanding in the development of moral cognition and moral action. *New directions for child and adolescent development*, *103*, 37–49.

Beck, S. R. (2015). Why what is counterfactual really matters: A response to Weisberg and Gopnik (2013). *Cognitive Science*, *40*(1), 253–256.

Beck, S. R., Robinson, E. J., Carroll, D. J., & Apperly, I. A. (2006, Mar). Children's thinking about counterfactuals and future hypotheticals as possibilities. *Child Development*, *77*(2), 413–426.

Bridgers, S., Jara-Ettinger, J., & Gweon, H. (2020). Young children consider the expected utility of others' learning to decide what to teach. *Nature Human Behaviour*, *4*(2), 144–152.

Byrne, R. M. (2016). Counterfactual thought. *Annual Review of Psychology*, *67*, 135–157.

Cushman, F. (2008). Crime and punishment: Distinguishing the roles of causal and intentional analyses in moral judgment. *Cognition*, *108*(2), 353–380.

Cushman, F., Sheketoff, R., Wharton, S., & Carey, S. (2013). The development of intent-based moral judgment. *Cognition*, *127*(1), 6–21.

Gerstenberg, T., Peterson, M. F., Goodman, N. D., Lagnado, D. A., & Tenenbaum, J. B. (2017, oct). Eye-tracking causality. *Psychological Science*, *28*(12), 1731–1744.

Gopnik, A., Sobel, D., Schulz, L., & 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*(5), 620–629.

Guajardo, N. R., & Turley-Ames, K. J. (2004). Preschoolers' generation of different types of counterfactual statements and theory of mind understanding. *Cognitive Development*, *19*(1), 53–80.

Hamlin, J. K., Wynn, K., & Bloom, P. (2007). Social evaluation by preverbal infants. *Nature*, *450*(7169), 557–559.

Harris, P. (1997). On realizing what might have happened instead. *Polish Quarterly of Developmental Psychology*, *3*, 161–176.

Harris, P. L., German, T., & Mills, P. (1996). Children's use of counterfactual thinking in causal reasoning. *Cognition*, *61*(3), 233–259.

Kleiman-Weiner, M., Gerstenberg, T., Levine, S., & Tenenbaum, J. B. (2015). Inference of intention and permissibility in moral decision making. In *Cogsci.*

Kominsky, J. F., Gerstenberg, T., Pelz, M., Singmann, H., Sheskin, M., & Keil, F. (2019). The trajectory of counterfactual simulation in development. In A. Goel, C. Seifert, & C. Freksa (Eds.), *Proceedings of the 41st Annual Conference of the Cognitive Science Society* (pp. 2044–2050). Montreal, QB: Cognitive Science Society.

Kuczaj, S. A., & Daly, M. J. (1979). The development of hypothetical reference in the speech of young children. *Journal of Child Language*, *6*(3), 563–579.

Kushnir, T., & Gopnik, A. (2005). Young children infer causal strength from probabilities and interventions. *Psychological Science*.

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

Lewis, D. (1973). Causation. *The Journal of Philosophy*, *70*(17), 556–567.

Martin, A., & Olson, K. R. (2013). When kids know better: Paternalistic helping in 3-Year-Old children. *Developmental Psychology*, *49*(11), 2071.

McCormack, T., Ho, M., Gribben, C., O'Connor, E., & Hoerl, C. (2018, jan). The development of counterfactual reasoning about doubly-determined events. *Cognitive Development*, *45*, 1–9.

Nelson, S. A. (1980). Factors influencing young children's use of motives and outcomes as moral criteria. *Child Development*, *51*(3), 823–829.

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

Nyhout, A., Henke, L., & Ganea, P. A. (2017). Children's counterfactual reasoning about causally overdetermined events. *Child Development*.

Rafetseder, E., & Perner, J. (2018, April). Belief and counterfactuality: A teleological theory of belief attribution. *Zeitschrift für Psychologie*, *226*(2), 110-121.

Rafetseder, E., Schwitalla, M., & Perner, J. (2013, March). Counterfactual reasoning: From childhood to adulthood. *Journal of Experimental Child Psychology*, *114*(3), 389–404.

Rai, T. S., & Fiske, A. P. (2011). Moral psychology is relationship regulation: Moral motives for unity, hierarchy, equality, and proportionality. *Psychological Review*, *118*(1), 57–75.

Schächtele, S., Gerstenberg, T., & Lagnado, D. A. (2011). Beyond outcomes: The influence of intentions and deception. In L. Carlson, C. Hölscher, & T. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 1860–1865). Austin, TX: Cognitive Science Society.

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.

Weisberg, D. S., & Gopnik, A. (2013, Aug). Pretense, counterfactuals, and bayesian causal models: Why what is not real really matters. *Cognitive Science*, *37*(7), 1368–1381.

Young, L., Cushman, F., Hauser, M., & Saxe, R. (2007). The neural basis of the interaction between theory of mind and moral judgment. *Proceedings of the National Academy of Sciences*, *104*(20), 8235–8240.

Yuill, N., & Perner, J. (1988). Intentionality and knowledge in children's judgments of actor's responsibility and recipient's emotional reaction. *Developmental Psychology*, *24*(3), 358–365.