Young children use mental simulation to reason about their performance

Peter Zhu, Chloe Chang, Tobias Gerstenberg, & Hyowon Gweon
`{pgzhu, changrc, gerstenberg, gweon}@stanford.edu`
Department of Psychology, Stanford University
Stanford, CA, USA

Abstract

Young children can use observed performance outcomes (e.g., successes and failures) to decide when to persist and which tasks to pursue. However, learners often face situations where outcomes either cannot be observed or are uninformative. Using a simple tablet game, we asked whether preschool-aged children can use mentally simulated outcomes to guide their decisions. In Experiment 1, the game froze mid-trial so the final outcome was unavailable. Children preferred to repeat the same game when their attempt would have resulted in success (versus failure). In Experiment 2, an on-screen agent intervened mid-trial, rendering the outcome uninformative about the child's performance. Children preferred to play the game without (versus with) the agent when their attempts would have been successful without the agent's intervention. These findings suggest that children can simulate alternative outcomes of their actions and use them to guide how they pursue future tasks.

Keywords: cognitive development; mental simulation; task choice; performance outcomes

Introduction

Imagine a child playing soccer in her local league. She might use a variety of signals to gauge her ability as a soccer player: how many goals she has attempted and scored, how many passes she has completed, and how easily she can avoid opponents' tackles. These performance outcomes—directly observable results of her own actions—provide an essential source of information for building an accurate representation of her abilities. In turn, these outcomes can guide further decisions about whether to keep practicing, apply for a more advanced league, or switch to a different sport.

Prior research suggests that children can leverage their own and others' performance outcomes (e.g., past successes and failures) to guide their subsequent actions. For example, even infants use information about others' successes and failures to identify the causes of their own failures (Gweon & Schulz, 2011) and calibrate their own effort on the task (e.g., Leonard et al., 2017; Lucca et al., 2020). By age two, children use the temporal pattern of their own successes and failures (e.g., improving performance) to infer their abilities (Zhu et al., 2024); by age five, children generate accurate predictions about how their performance outcomes will change over time (Zhang et al., 2025) and use the gradual change in their observed performance to decide when to persist on a task (Leonard et al., 2023). These findings suggest that children consider the outcomes of their own actions as diagnostic of their abilities.

Clear performance outcomes, however, are not always available; learners often face situations where they must go beyond what they directly observe. For instance, learners may have initiated an attempt but the outcome is unavailable, either due to an unexpected interruption or external interference. If a soccer ball was headed toward the goal but a referee catches the ball mid-flight and pauses play, the player never saw whether the shot would have succeeded. Alternatively, outcomes may be available but uninformative—for example, a teammate may accidentally deflect the ball on its way to the goal, causing it to miss. In this case, while there is an observed outcome (miss), it may be uninformative about the player's true competence.

In these situations, we as adults cannot help but think about *what could have happened*: that the ball was heading for the goal and likely would have made it if the interruption had not occurred. Generating these alternative outcomes, however, requires going beyond what is directly observable; it requires using available evidence to mentally simulate an outcome that could have occurred but never actually did. The current work investigates whether young children can go beyond observed evidence to reason about their performance outcomes. Can they spontaneously engage in mental simulation to generate plausible outcomes of their own actions?

A longstanding line of work suggests that mental simulation can support learning and reasoning by generating outcomes "in the mind" (e.g., Lombrozo, 2024). Humans deploy simulation in a wide variety of settings, such as mechanical and physical reasoning (e.g., Dasgupta et al., 2018; Hegarty, 2004), physical scene judgments (e.g., Battaglia et al., 2013; Gerstenberg et al., 2017), social cognition (e.g., Wu et al., 2023), spatial reasoning (e.g., Shepard & Metzler, 1971), and scientific discovery (e.g., Brown, 1986). By using their mental models of the world, humans not only predict what might happen but also infer what might have happened counterfactually (Gerstenberg, 2024). By simulating possible scenarios and event outcomes, adults can go beyond "learning-by-doing" and benefit from "learning-by-thinking".

The ability to engage in mental simulation—at least in relatively rudimentary forms—may emerge early in life. Infants move their eyes to anticipate objects moving in space (even when the objects themselves are occluded from view; Johnson et al., 2003), and coordinate their own reaching behavior to act upon objects already in motion (e.g., von Hofsten et al., 1998), suggesting that infants can use their mental model of the physical world to generate online predictions about object trajectories. By age two, children can use their knowledge about the world in order to construct imaginary worlds in the context of pretense and pretend play, such as predicting that a character's head would be wet after imaginary tea was poured on it (Harris et al., 1993).

Research has also found a later-emerging capacity to engage in more complex mental simulations. The ability to entertain multiple simulated possibilities appears to emerge after four years of age (Leahy and Carey, 2020; Redshaw and Suddendorf, 2020; but see also Cesana-Arlotti et al., 2025), and so does the ability to reason about hypothetical outcomes (i.e., simulating a possible future outcome; e.g., Nyhout et al., 2023) and counterfactual outcomes (i.e., simulating outcomes that could have happened if something had been different; e.g., Nyhout and Ganea, 2019; Rose et al., 2025; but see also Ong et al., 2021; Rafetseder et al., 2013, for development beyond the preschool years). In one study (Kominsky et al., 2021), children were asked how a soccer ball would have moved if a wooden block had not altered its path; while 4- and 5-year-olds used mental simulation, their answers systematically differed from those of adults and older children. Together, this body of work shows substantial changes in 4- to 5-year-olds' capacity for mental simulation, making this a particularly interesting window to study its development.

Notably, this past work has primarily focused on children's ability to use mental simulation as third-party observers, raising questions about whether they can also simulate the outcomes of their own actions. One possibility is that the same capacity for mental simulation underlies children's inferences as third-party observers and first-person actors, supporting similar developmental trajectories. On the other hand, a large body of research has suggested that children make overwhelmingly optimistic predictions about how well they will do on a range of tasks—memory, motor, and probability-based (see Leonard & Sommerville, 2025; Xia et al., 2024, for reviews)—raising the possibility that such optimism may affect how they simulate their own performance outcomes. That is, children may be similarly biased in reasoning about simulated outcomes that they *could have* produced, entertaining optimistic alternatives even when they are implausible. Finally, it is also possible that young children rely exclusively on observed outcomes to reason about their own performance; even though simulated outcomes can be an informative learning signal, children's limited representational capacity may prevent them from utilizing simulated outcomes, especially when these outcomes differ from observed outcomes.

Here, we report two studies that investigate whether children use mental simulation to reason about their performance. Drawing inspiration from prior work (e.g., Kominsky et al., 2021), we developed a virtual soccer game (see Figure 1): a billiards-style task in which a ball moves towards a goal on a screen. Critically, rather than passively viewing the event, participants themselves attempted (i.e., by blowing into the hole on a tablet device) to launch a ball towards a goal, such that the outcome encoded the consequence of their actions.

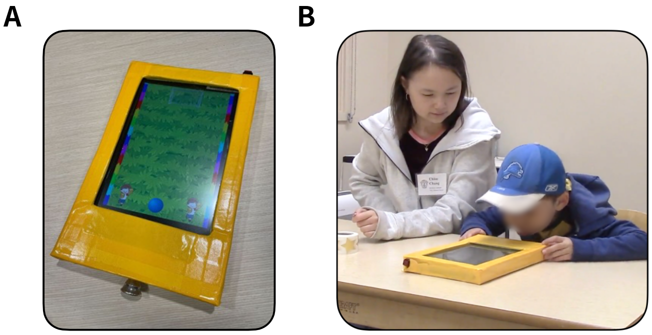

Figure 1: **Experiment set-up**. (A) Experimental device; an iPad placed within a box, with a metal trumpet mouthpiece at the bottom. (B) A participant blowing into the mouthpiece, which appeared to launch the blue ball towards the goal (in reality, the trajectory of the ball was predetermined).

Experiment 1: Direct Outcome Unavailable

Experiment 1 investigated whether young children can mentally simulate their performance outcomes when the actual outcome is unavailable. In the virtual soccer game described above (see Figure 1), children blew into the bottom of a box to launch a ball towards a goal. On the critical test trial, the game froze mid-trial, rendering the final outcome unavailable. The key difference between conditions was that the ball was headed either straight towards the goal ('Almost condition') or veering for a miss ('Miss condition') before the game froze. If children can use mental simulation to reason about their performance, they should be more likely to play another round of the same game (rather than switch to another task) in the Almost condition than in the Miss condition.

We preregistered a sequential Bayes Factor (BF) analysis to iteratively evaluate the effect of interest until a predetermined evidentiary criterion or the final N was met (e.g., Mani et al., 2021; Schönbrodt et al., 2017). This provides a data-driven approach to determine the final sample size; researchers predetermine their maximum sample size but also have the option to stop data collection early in the presence of strong evidence for either the null or the predicted hypothesis.[1]

Methods

Participants Forty-four (n = 22 per condition) 4- and 5-year-old children were tested at a local preschool (M_{age} = 4.88, SD_{age} = 0.42). We excluded five additional participants for the following preregistered exclusion criteria: not completing the study (n = 3), experimenter error (n = 1), and tech error (n = 1). See Results for sequential sampling procedure to determine the final sample size.

Materials Children performed the task on an iPad device housed in a casing made of foam board and yellow duct tape.

---

[1] Data, analysis code, stimuli, study scripts, and preregistrations for both experiments are available at https://github.com/pzhu222/soccr_cogsci_2026.

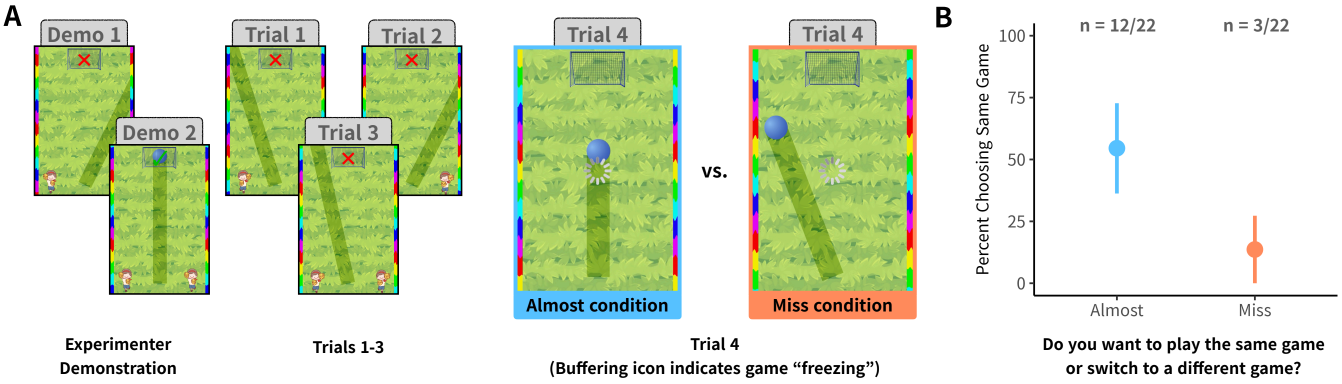

Figure 2: **Experiment 1 methods and results**. (A) Final outcomes of each trial: dark lines indicate the trajectory the ball took (the lines were visible to the participants as the ball traveled along its path). After seeing the experimenter fail (Demo 1) and succeed (Demo 2), participants attempted the game and failed three times (Trials 1-3). On the fourth trial, the ball froze midway, but its trajectory before freezing differed by condition (final distance to the goal matched across conditions). (B) Proportion of children choosing to play the same game (vs. switch to a different task). Error bars show 95% bootstrapped CIs.

A red button and a trumpet mouthpiece were attached to the casing (see Figure 1). Video stimuli, presented on the iPad, were constructed using Keynote. Finally, a roll of gold star stickers was used during the study.

Procedure All children were tested in a quiet room at their preschool. Children were seated at a long table, with the experimenter sitting beside them. The experimenter told children that they would play a game and pressed the red button on the top of the yellow iPad casing which appeared to "begin" the game (in reality, all aspects of the game were surreptitiously controlled by the experimenter). With the initial scene on the screen, the experimenter said: "You have to make this ball (at the bottom of the screen) go into this goal here (at the top of the screen). In order to make the ball go, you have to blow really hard right here (pointing to the mouthpiece at the bottom of the casing). If you make the ball go into the goal, you get one of these cool stickers!"

The experimenter demonstrated the game by blowing into the mouthpiece and missing the goal, which displayed a red "X" and made a loud buzzer sound. The experimenter then attempted the task a second time, pressing the red button between trials to give the appearance of "resetting" the game, and succeeded on their second attempt (displaying a green check and playing a ding sound effect). After succeeding, the experimenter gave themselves a sticker. Children then attempted the game three times, missing the goal each time.

After the child's third attempt, the experimenter offered the child one more attempt (and reminded them that they'd receive a star sticker if successful). Critically, the trajectory of their ball differed by condition. In the Almost condition, the ball headed straight towards the goal, whereas in the Miss condition, the ball veered for a miss. Prior to the ball reaching its final outcome, the game appeared to freeze, pausing the ball at its current location, stopping the music, and displaying a rotating gray circle (as if the game was buffering). We controlled for the low-level properties of the final scene across conditions: the total time the ball spent in motion until the game froze and the Euclidean distance between the ball and the goal in the final scene. See Figure 2A.

When the game froze, the experimenter noted: "Well, it looks like I will have to restart this game . . . " Then the experimenter asked the key test question: "If you want to keep playing this game, I'll just have to restart my game—it will be quick but you'll have to wait. Or, we can play a different game, and you can do that right now.[2] So, do you want to wait while I restart this game or do you want to play a different game right now?" Our key measure was children's decision to play the same game or switch to a different game.

Results

Across conditions, children's observed performance outcomes in the first three trials were identical (all misses) and the fourth outcome was unavailable. If children can use simulated outcomes to make strategic task decisions, those children in the Almost condition (whose simulated outcome was a success) would be more likely to play the same game than children in the Miss condition (whose simulated outcome was a miss). Importantly, our key hypothesis for the current study focused on a condition difference rather than a difference compared to chance (50%), since children may have had a baseline preference to keep playing the same game (e.g., the star sticker may be especially appealing) or to switch tasks (e.g., repeated failure may feel especially discouraging).

Results are in Figure 2B. To test our predictions, we ran a Bayesian generalized linear model with default weakly informative priors in the `rstanarm` package (Goodrich et al., 2025). We computed BFs using the `bridgesampling` package (Gronau et al., 2020), comparing a model with a condi-

---

[2] It was clear in context that restarting the game meant starting a fresh trial, rather than simply unfreezing the current trial.

tion term (i.e., `choice ~ condition`) and one without (i.e., intercept-only model; `choice` ~ 1). Our Bayesian sequential sampling procedure involved testing an initial sample of n = 40 (20 per condition), then evaluating the BF on the hypothesis of a condition effect after every 4 participants. Stopping criteria (based on guidelines for reporting Bayesian statistics; e.g., Schönbrodt et al., 2017) were set at a $BF_{10} > 10$ (i.e., strong evidence in favor of the hypothesis of a condition effect), or a $BF_{01} > 5$ (i.e., strong evidence in favor of the null hypothesis), or a max of $n = 72$ (36 per condition).

We were able to terminate data collection early given strong evidence in favor of the hypothesis, resulting in a final sample of $n = 44$. As predicted, children in the Almost condition were more likely to want to keep playing the same game ($n = 12/22$) compared to children in the Miss condition ($n = 3/22$, $BF_{10} = 11.33$, $\beta = -2.09$, 95% CrI = [−3.70, −0.74]). In exploratory analyses, we did not find evidence that the effect of condition differed by age ($BF_{10} = 0.67$, $\beta_{\text{condition:age}} = 0.004$, 95% CrI = [−1.35, 1.38]).[3]

Discussion

Experiment 1 provides evidence that children can go beyond what is directly observable. Note that performance outcomes were identical across conditions, and the only difference was the partial trajectory of the ball in the final trial (in the absence of a direct outcome), meaning that children could not simply integrate past outcomes to guide their decision. Yet, children were able to generate a simulated performance outcome to guide their decision of which game to play next.

The experimental procedure and the final choice—stay or switch—were deliberately designed to present an ambiguous choice. Children had reasons to stay (e.g., win a sticker) but also to switch (e.g., they had missed 3 times, and would need to wait to restart the current game). Our results suggest that the simulated outcome pushed children's decisions in opposite directions. Children in the Almost condition may have taken the simulated success as indicative of their ability to succeed again, whereas children in the Miss condition may have taken their simulated failure as evidence that they would fail.

While these findings provide initial support for children's ability to use mental simulation to reason about their own performance outcomes, Experiment 1 leaves open a few important questions. First, the trajectory of the ball in the final, fourth trial differed across conditions, raising the possibility that children's responses were driven by low-level cues. For instance, children may have formed an association between straight path and success based on the experimenter's attempts, or children in the Almost condition may have been encouraged by the novelty of their final trajectory (straight path) compared to those in the Miss condition who always saw a diagonal path.

Second, in Experiment 1, children simply had to generate and maintain a single representation—the predicted location of the ball, extrapolated from its past trajectory—in the absence of any competing alternatives. Indeed, such simple predictions about physical events have been demonstrated even in infants' anticipatory looks (Johnson et al., 2003). While such "forward projection" still requires simulation, our findings raise questions about whether children can use simulated outcomes even in the presence of an observable outcome, especially when it is uninformative. For instance, if a child's ball was headed towards the goal but an external agent intervened, causing a miss, would children use simulated outcomes (rather than the actual outcome) to decide what to do next?

It is possible that children may prioritize directly observed outcomes over simulated evidence. Given that concrete outcomes are arguably more salient than simulated outcomes, children may fail to suppress a prepotent response—based on the actual observed outcome—instead of using the simulated outcome (see Carlson & Moses, 2001, for an example in the context of Theory of Mind). Furthermore, if children consider the observed outcome as an immutable aspect of the scene that cannot be changed or revised (see Kominsky et al., 2021, for a discussion), children may have difficulty considering simulating outcomes that run contrary to the direct outcome.

Experiment 2 was designed with a few goals in mind. First, we wanted to address alternative explanations that rely on superficial, low-level differences in the ball's trajectory. Second, we aimed to explore children's ability to use simulated performance when direct outcomes are available but uninformative. Finally, going beyond the decision to switch vs. stay, Experiment 2 implemented a socially oriented task decision: Whether to perform the task with the help of another agent.

Experiment 2: Direct Outcome Uninformative

In Experiment 2, children played the same soccer game. The key difference was the presence of a character called a "Gazorp", who intervened mid-trial to alter the trajectory of the ball, rendering the final outcome uninformative about the child's competence. Instead of varying the trajectory of the ball, we varied the location of the goal: the child would have made the goal if the Gazorp had not intervened ('Hinder condition') or would have missed the goal regardless ('Help condition'). This allowed us to rule out low-level alternatives in Experiment 1, such as the novelty in observed paths.

Method

Participants We used the same sequential sampling procedure as in Experiment 1. The final sample for Experiment 2 consisted of sixty-four (n = 32 per condition) 4- and 5-year-old children tested at a museum ($M_{age} = 4.99$, $SD_{age} = 0.58$). We excluded an additional twenty participants for an ambiguous or uncodeable choice (n = 4), fussing out or not completing the study (n = 11), interference (e.g., parents, siblings; n = 2), experimenter error (n = 1), having a parent-reported developmental delay (n = 1), and tech error (n = 1).

---

[3] To assess whether an effect of condition differed across age, we computed the BF by comparing a model containing only the main effects of condition and age to a model that contained both main effects as well as their interaction.

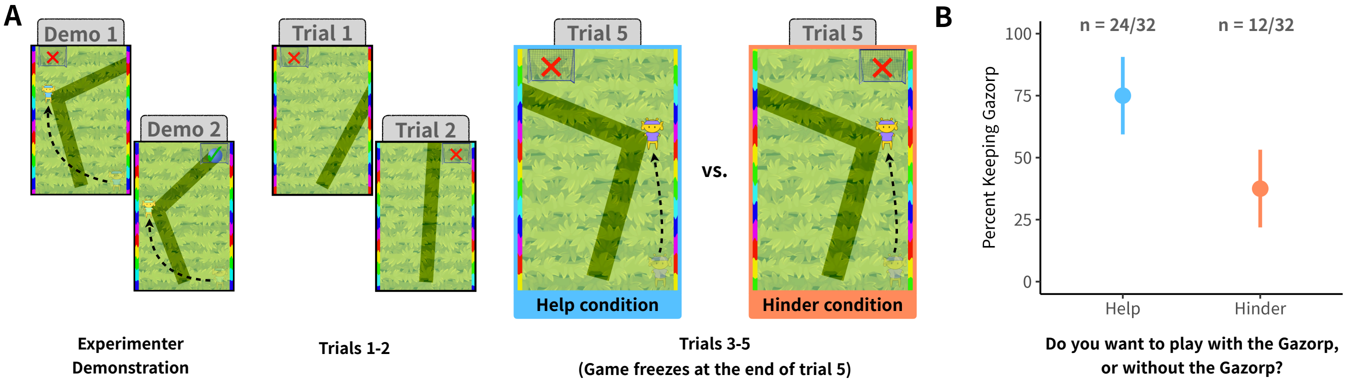

Figure 3: **Experiment 2 methods and results**. (A) A game character ('Gazorp') appeared on the bottom right corner of the screen. Images indicate final outcomes of each trial; low-opacity Gazorp depicts its starting position, with an arrow indicating the Gazorp's motion to its final position (participants did not see the arrow). The experimenter demonstrated the game with a blue and green Gazorp, first missing the goal (Demo 1), then making the goal (Demo 2). Participants then attempted the game twice without a Gazorp (missing each time; Trials 1-2). On Trials 3-5, children attempted the game with a purple Gazorp present: each attempt resulted in a miss but the purple Gazorp either appeared to help or hinder the child's attempts. Image shows only Trial 5: Trial 3 is nearly identical and Trial 4 is a mirror image. (B) Proportion of children choosing to play with (vs. without) the Gazorp. Error bars show 95% bootstrapped CIs.

Procedure Stimuli and procedure were nearly identical to those of Experiment 1, except for minor stylistic changes. One key difference from Experiment 1 was the placement of the goal—instead of being in the middle, it appeared in the top left or the top right corner of the screen depending on the trial.

The experimenter introduced Gazorp to the child by saying that "some Gazorps are nice, and other Gazorps are mean, and we don't know whether this Gazorp with the blue shirt is a nice one or a mean one." The experimenter then demonstrated the game. First, their ball was headed straight for the goal (top left) before the blue Gazorp appeared to hit the ball off track (see Figure 3). Next, as the ball was headed for a miss, a green Gazorp moved to hit the ball into the goal (top right), resulting in a success.

Then, children tried the game twice without a Gazorp present, missing each time. On the third trial, a purple Gazorp appeared in the bottom right corner. The experimenter reminded children that some Gazorps are nice and others are mean, and that it is unknown whether the purple Gazorp is nice or mean. Children then attempted the game three more times with the purple Gazorp. During each of these attempts, the Gazorp intervened on the trajectory of the ball midway through its path, causing it to ultimately miss the goal.

The key condition difference was the position of the goal in trials 3–5: in the Help condition, the Gazorp intervened on the ball when it veered for a miss (Figure 3), whereas in the Hinder condition, the intervention occurred when the ball headed straight for the goal. While the actual trajectory that children observed across the two conditions was exactly the same, what varied across conditions was the position of the goal: either out of, or in, the path of the ball's motion prior to the Gazorp's intervention. Across both conditions, the ball always missed the goal, such that the direct outcomes that children observed were also held fixed across conditions. Children saw three instances of the helping or hindering action by the same purple Gazorp; across successive trials, the trajectory and goal position were flipped left to right. After the final outcome in the fifth trial, the game "froze," displaying the rotating icon as in Experiment 1.

The experimenter then offered to restart the game, brought out a roll of stickers, and told children that they would receive a sticker if they successfully made it into the goal on their next attempt. Finally, children were asked whether they wanted to play with or without the Gazorp, with the experimenter saying: "If it's a nice Gazorp, maybe it will help you make it, but if it's a mean Gazorp, maybe it will try to make you miss. So, do you want to play with the Gazorp or without the Gazorp?" Our key measure was children's decision to either play with or without the Gazorp; while children may engage in social inference (i.e., "is the Gazorp a helper or hinderer?") to guide their decisions, our stimuli were designed such that simulation would be required to support such an inference.

Results

Results are depicted in Figure 3B. Like Experiment 1, our hypothesis focused on a difference across conditions, rather than chance. The analyses were identical to Experiment 1, computing a BF comparing a model with a condition term and one without. We ended data collection early given strong evidence in favor of the hypothesis, leading to a final sample of $n = 64$. We found that children in the Help condition were more likely to want to keep the Gazorp ($n = 24/32$) compared to children in the Hinder condition ($n = 12/32$, $\mathrm{BF}_{10} = 12.41$, $\beta = -1.63$, 95% CrI = $[-2.75, -0.59]$). In exploratory analyses, we did not find any effect of age ($\mathrm{BF}_{10} = 0.61$, $\beta_{\text{condition:age}} = -0.15$, 95% CrI = $[-1.24, 0.96]$).

Discussion

Experiment 2 results suggest that children were able to discount a direct, observed outcome that is uninformative about their ability; instead, they used a simulated outcome in order to decide what to do next. Since the exact trajectories and outcomes were matched across conditions, our task was robust against a number of low-level alternatives; for example, children could not simply track the covariance between the outcome and the agent, or use differences in the paths across conditions. While the only difference between the two conditions was the position of the goal, this was enough to modulate children's decisions about whether to play with the agent.

One may wonder if the results could still be explained by a low-level cue: the distance between the goal and the ball's exit point. Since this distance was longer in the Hinder condition, children may have used this—rather than the simulated outcome—to infer the Gazorp's helpfulness and make their final decision. Thus, we are running a preregistered follow-up in which the external intervention (experimenter, rather than Gazorp) leads to success—controlling for the final position of the ball—and children are asked to assign credit to themselves or the experimenter. Preliminary results suggest that children are more likely to indicate themselves as doing better when their initial action could have also produced a successful goal. These data would suggest that children can use simulated performance to determine who deserves credit when evidence is confounded, in addition to showing that the distance between the ball and the goal cannot alone explain their task decisions.

General Discussion

Across two experiments, we examined preschool-aged children's ability to use mental simulation to reason about their performance outcomes. We show that children are able to use simulated performance outcomes to guide their future task choices when direct, observable outcomes are either unavailable (Experiment 1) or uninformative (Experiment 2).

Our findings are notable in light of prior work highlighting children's unwavering optimism bias (Xia et al., 2024). In the current task, we did not find evidence that such a bias influenced the ways children simulated their own performance outcomes. However, our findings raise an interesting question of how, and when, children's simulations for the outcomes of their own performance may differ from their simulations of third-party events. One possibility is that children may show *earlier* competence for accurate simulations about their own actions (Goddu & Gopnik, 2024); it may be especially motivating to produce accurate information about the self or it may be easier to simulate the outcomes of one's own actions. Alternatively, it may be the case that an optimism bias may only emerge when directly compared against simulations for external events or another agent's actions. As the current task did not directly contrast children's simulations for the self vs. the world, this remains an open question for future work.

Our stimuli drew from past work studying counterfactual simulation (e.g., Gerstenberg et al., 2017; Kominsky et al., 2021): a billiards-style physical scene in which a ball moves towards a target. These paradigms have been particularly useful in studying whether people engage in physical simulation to generate alternative counterfactual outcomes. While it remains possible that children in our study did engage in counterfactual reasoning (e.g., "If the game had not frozen, I would have made it!", or "If the Gazorp had not hit the ball, I would have made it!"), it was not necessary to succeed in our task. Children may have simply predicted the goal's path forward before the game froze or before Gazorp intervened and used it to make their decision. Both accounts are consistent with the use of simulation for one's own performance, but further work can tease apart these possibilities in order to understand the exact representations that guide task choice.

Our dependent measures across studies highlight different ways in which simulation can guide task choice: Should I stay on task or switch (Exp. 1)? Should I keep or drop a component of the task (Exp. 2)? Who deserves credit for a successful outcome (ongoing Exp. 3)? These "indirect" measures require using simulated outcomes for a task-based decision, and were used in part to reduce the demands of directly asking about simulated outcomes (see Rose et al., 2025, for a discussion in counterfactual language). Thus, our experiments show that mental simulation can guide children's downstream behaviors and decisions. More work is needed, however, to better understand the limitations of simulated outcomes. Are they just as good as (or almost as good as) actual outcomes or do they have weaker evidentiary value? When do we "trust" our simulations, and when do we think they may be miscalibrated? These questions are worth exploring not only in the domain of predicting performance outcomes, but in people's ability to use simulated outcomes more generally.

Importantly, one key use for simulation is in cases in which a learner may need to simulate *from scratch*, without having taken an action at all. While the current work focuses on cases in which an action has already occurred, learners may also rely on simulation in order to estimate the difficulty of novel tasks (Gweon et al., 2017) or to generate performance outcomes when taking an action may be risky or costly. While a body of work in sports and music has highlighted the role of "mental practice" in aiding learning (especially when physical practice is unavailable or costly; e.g., Coffman, 1990; Lindsay et al., 2019; Zecker, 1982), an account of how, and when, learners simulate the outcomes of their actions before they have happened still remains an open question.

One of the most important obstacles in early childhood is figuring out which challenges to pursue, when to persist, and when to switch to a new task. Building on a body of work demonstrating children's abilities to reason about observed performance outcomes, here we show that children spontaneously use mental simulation to generate performance outcomes. These outcomes can guide their future task decisions and even help them learn about their competence. This ability—learning-by-thinking about the self—may serve as a key tool for learning throughout childhood and beyond.

Acknowledgments

We thank the Social Learning Lab and Causality in Cognition Lab for helpful comments and feedback. We also thank Bing Nursery School and the Palo Alto Junior Museum and Zoo; thank you to Chia-wa Yeh for help recruiting and supervising testing, and to the families and children who participated. This work was supported by grants from the Stanford Institute for Human-Centered Artificial Intelligence and the Cooperative AI Foundation (TG), and James S. McDonnell Foundation and NSF (HG).

References

Battaglia, P. W., Hamrick, J. B., & Tenenbaum, J. B. (2013). Simulation as an engine of physical scene understanding. *Proceedings of the National Academy of Sciences*, *110*(45), 18327–18332.

Brown, J. R. (1986). Thought experiments since the scientific revolution. *International Studies in the Philosophy of Science*, *1*(1), 1–15.

Carlson, S. M., & Moses, L. J. (2001). Individual Differences in Inhibitory Control and Children's Theory of Mind. *Child Development*, *72*(4), 1032–1053.

Cesana-Arlotti, N., Jáuregui, S., Mazalik, P., Nichols, S., & Halberda, J. (2025). Logical Concepts of (Im)possibility Guide Young Children's Decision-Making. *Developmental Science*, *28*(5), e70044.

Coffman, D. D. (1990). Effects of Mental Practice, Physical Practice, and Knowledge of Results on Piano Performance. *Journal of Research in Music Education*, *38*(3), 187–196.

Dasgupta, I., Smith, K. A., Schulz, E., Tenenbaum, J. B., & Gershman, S. J. (2018, May). Learning to act by integrating mental simulations and physical experiments.

Gerstenberg, T. (2024). Counterfactual simulation in causal cognition. *Trends in Cognitive Sciences*.

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

Goddu, M. K., & Gopnik, A. (2024). The development of human causal learning and reasoning. *Nature Reviews Psychology*, *3*(5), 319–339.

Goodrich, B., Gabry, J., Ali, I., & Brilleman, S. (2025). Rstanarm: Bayesian applied regression modeling via Stan. [R package version 2.32.2].

Gronau, Q. F., Singmann, H., & Wagenmakers, E.-J. (2020). Bridgesampling: An R Package for Estimating Normalizing Constants. *Journal of Statistical Software*, *92*(10), 1–29.

Gweon, H., Asaba, M., & Bennett-Pierre, G. (2017). Reverse-engineering the process: Adults' and preschoolers' ability to infer the difficulty of novel tasks. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *39*.

Gweon, H., & Schulz, L. (2011). 16-Month-Olds Rationally Infer Causes of Failed Actions. *Science*, *332*(6037), 1524–1524.

Harris, P. L., Kavanaugh, R. D., Wellman, H. M., & Hickling, A. K. (1993). Young Children's Understanding of Pretense. *Monographs of the Society for Research in Child Development*, *58*(1), i–107.

Hegarty, M. (2004). Mechanical reasoning by mental simulation. *Trends in Cognitive Sciences*, *8*(6), 280–285.

Johnson, S. P., Amso, D., & Slemmer, J. A. (2003). Development of object concepts in infancy: Evidence for early learning in an eye-tracking paradigm. *Proceedings of the National Academy of Sciences*, *100*(18), 10568–10573.

Kominsky, J. F., Gerstenberg, T., Pelz, M., Sheskin, M., Singmann, H., Schulz, L., & Keil, F. C. (2021). The trajectory of counterfactual simulation in development. *Developmental Psychology*, *57*(2), 253–268.

Leahy, B., & Carey, S. E. (2020). The Acquisition of Modal Concepts. *Trends in Cognitive Sciences*, *24*(1), 65–78.

Leonard, J. A., Cordrey, S. R., Liu, H. Z., & Mackey, A. P. (2023). Young children calibrate effort based on the trajectory of their performance. *Developmental Psychology*, *59*(3), 609–619.

Leonard, J. A., Lee, Y., & Schulz, L. E. (2017). Infants make more attempts to achieve a goal when they see adults persist. *Science*, *357*(6357), 1290–1294.

Leonard, J. A., & Sommerville, J. A. (2025). A unified account of why optimism declines in childhood. *Nature Reviews Psychology*, *4*(1), 35–48.

Lindsay, R., Spittle, M., & Larkin, P. (2019). The effect of mental imagery on skill performance in sport: A systematic review. *Journal of Science and Medicine in Sport*, *22*, S92.

Lombrozo, T. (2024). Learning by thinking in natural and artificial minds. *Trends in Cognitive Sciences*, *28*(11), 1011–1022.

Lucca, K., Horton, R., & Sommerville, J. A. (2020). Infants rationally decide when and how to deploy effort. *Nature Human Behaviour*, *4*(4), 372–379.

Mani, N., Schreiner, M. S., Brase, J., Köhler, K., Strassen, K., Postin, D., & Schultze, T. (2021). Sequential Bayes Factor designs in developmental research: Studies on early word learning. *Developmental Science*, *24*(4), e13097.

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

Nyhout, A., Sweatman, H., & Ganea, P. A. (2023). Children's hypothetical reasoning about complex and dynamic systems. *Child Development*, *94*(5), 1340–1355.

Ong, D., Asaba, M., Lim, H. Y., Chen, P., & Gweon, H. (2021). "If only Santa had one more present": Exploring the development of near-miss counterfactual reasoning. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *43*.

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

Redshaw, J., & Suddendorf, T. (2020). Temporal Junctures in the Mind. *Trends in Cognitive Sciences*, *24*(1), 52–64.

Rose, D., Zhang, S., Bridgers, S. E. C., Gweon, H., & Gerstenberg, T. (2025, October). Probing counterfactual thinking without counterfactual language.

Schönbrodt, F. D., Wagenmakers, E.-J., Zehetleitner, M., & Perugini, M. (2017). Sequential hypothesis testing with Bayes factors: Efficiently testing mean differences. *Psychological Methods*, *22*(2), 322–339.

Shepard, R. N., & Metzler, J. (1971). Mental Rotation of Three-Dimensional Objects. *Science*, *171*(3972), 701–703.

von Hofsten, C., Vishton, P., Spelke, E. S., Feng, Q., & Rosander, K. (1998). Predictive action in infancy: Tracking and reaching for moving objects. *Cognition*, *67*(3), 255–285.

Wu, S. A., Sridhar, S., & Gerstenberg, T. (2023). A computational model of responsibility judgments from counterfactual simulations and intention inferences. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *45*.

Xia, M., Poorthuis, A. M. G., & Thomaes, S. (2024). Children's overestimation of performance across age, task, and historical time: A meta-analysis. *Child Development*, *95*(3), 1001–1022.

Zecker, S. G. (1982). Mental Practice and Knowledge of Results in the Learning of a Perceptual Motor Skill. *Journal of Sport and Exercise Psychology*, *4*(1), 52–63.

Zhang, X., Carrillo, B. A., Christakis, A., & Leonard, J. A. (2025). Children Predict Improvement on Novel Skill Learning Tasks. *Child Development*, *96*(3), 1177–1188.

Zhu, P., Aranda, V., Keene, G. E., & Gweon, H. (2024). Two-year-olds can reason about the temporal structure of their performance. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *46*.