Stop, children what's that sound? Multi-modal inference through mental simulation

Joseph Outa, Xi Jia Zhou, Hyowon Gweon, Tobias Gerstenberg
{joouta, xijiazho, gweon, gerstenberg}@stanford.edu
Department of Psychology, Stanford University, USA

Abstract

Human adults can figure out what happened by combining evidence from different sensory modalities, such as vision and sound. How does the ability to integrate multi-modal information develop in early childhood? Inspired by prior computational work and behavioral studies with adults, we examined 3- to 8-year-old children's ability to reason about the physical trajectory of a ball that was dropped into an occluded Plinko box. Children had to infer in which one of three holes the ball was dropped based on visual information (i.e., where the ball landed) and auditory information (i.e., the sounds of the ball colliding with parts of the box). We compare children's responses to the predictions of four computational models. The results suggest that although even the youngest children make systematic judgments rather than randomly guessing, children's ability to integrate visual and auditory evidence continues to develop into late childhood.

Keywords: mental simulation, intuitive physics, vision, audition, cross-modal integration, heuristics

Introduction

Humans are good at figuring out what happened. From some rocks on the ground, a geologist infers when the Ice Age began, and from a bullet hole in the wall, a forensic scientist figures out who committed the crime. Remarkably, the ability to reconstruct "what happened" is not limited to those with expert knowledge. Without degrees in geology or forensic science, people routinely use sparse evidence to draw rich inferences about the past. For example, adult participants can infer what path someone took based on the location of cookie crumbs on the floor (Lopez-Brau et al., 2020), or infer whether there is an object or an agent behind a curtain based on the pattern of sounds that were generated (Schachner & Kim, 2018; Kim & Schachner, 2021). The scope of such inferences extends beyond the physical world: from the surprised look on a friend's face, we can infer that something unexpected must have happened (Wu et al., 2021), and when one person gets blamed more than another, we get a sense for what each person must have done (Davis et al., 2021).

Recent computational work has examined how adult participants use their intuitive understanding of the physical world to infer what happened in the past (Smith & Vul, 2014; Gerstenberg et al., 2018), explain what happened in the present (Gerstenberg, Goodman, et al., 2021), predict what happens next (Smith & Vul, 2012; Battaglia et al., 2013), or plan to take actions that bring about desired outcomes (Allen et al., 2020). This work assumes that people's mental model of the physical world is similar to the kinds of physics engines that are used to generate physically realistic interactions in modern computer games (Ullman et al., 2017; Gerstenberg & Tenenbaum, 2017, but see Ludwin-Peery et al., 2021).

Critically, reconstructing the past often involves an integration of different cues that can span multiple modalities. Consider, for instance, the Plinko box at the top of Figure 1. Although you could guess in which hole the ball was dropped just based on where it landed in the sand, you could make a better guess if you had also heard the sounds it made when it was dropped. Gerstenberg, Siegel, & Tenenbaum (2021) tested adult participants' inferences in this task. In the 'vision' condition, participants only got to see the final location of the ball. In the 'vision & sound' condition, the box was first covered up and participants heard what sounds the ball made as it was dropped. The cover was then removed so that participants saw where the ball landed, and they were asked to figure out from which hole the ball was dropped. Adults were more accurate at figuring out in which hole the ball was dropped when they had access to both visual and auditory information rather than visual information only (Figure 1 bottom).

Despite recent advances in understanding the cognitive processes that support such integration of multiple sensory cues, little is known so far about how these inferences develop in early childhood. Recent proposals suggest that adults' ability to simulate physical events may be rooted in early-emerging knowledge about the physical world (Lake et al., 2017). Decades of developmental research have found that even infants have an intuitive, theory-like understanding of the physical world that may provide the foundations for these inferences (Spelke et al., 1992; Ullman & Tenenbaum, 2020). From this perspective, one might predict that the ability to integrate different cues to reason about physical events may also emerge relatively early in life.

However, existing work suggests that young children struggle with integrating multiple sources of information until late childhood. For instance, one study has examined children's ability to integrate visual and haptic information by having them discriminate the size and orientation of physical blocks (Gori et al., 2008). While 8- to 10-year-olds readily integrated the evidence in a statistically optimal fashion, weighting each modality appropriately based on its reliability, children younger than 8 years of age struggled to do so. Their

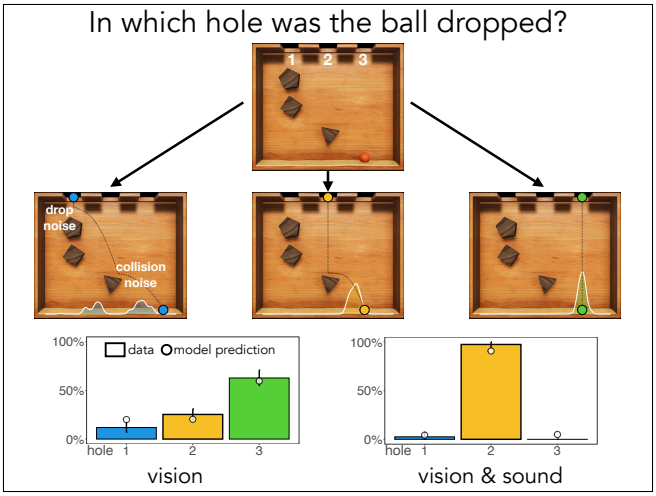


Figure 1: Illustration of the simulation model. To infer in which hole the ball was dropped, the model simulates where the ball would end up for each hole, and what sounds it would make along the way. The densities in the middle row summarize the outcomes of many simulation runs; the paths indicate the ground truth. The plots show adult participants' inferences (bars) and the model's posterior belief (points) based on visual information only (left) versus both visual and auditory information (right). On this trial, the ball was in fact dropped in the middle hole (hole 2).

responses were dominated either by vision or touch regardless of reliability. Yet, this study focused on online integration of visual and haptic information for making fine-grained perceptual distinctions (e.g., which object is taller?), rather than examining children's ability to reason about unobserved events.

More recent work provides some insights into young children's ability to reason about physical states of the world based on sensory information. For instance, 4- to 8-year-olds can use sound to discriminate between multiple hypotheses, and the duration of their exploration (i.e., shaking the box to figure out how many objects are inside) reflects the difficulty of teasing apart different hypotheses (Siegel et al., 2021). By 6 years of age, children can readily infer whether the water was hot or cold from the sound of the water being poured into a glass (Agrawal et al., 2020). These studies focused on whether children can use information from a single sensory modality (e.g., sounds), leaving open the question of how they might integrate multiple cues from different modalities.

In this paper, we adapt the Plinko task to study how children develop the ability to draw inferences about what happened based on visual and auditory information. This task offers a simple and intuitive way to assess causal inferences while allowing researchers to flexibly manipulate what sensory information is available. It also allows us to compare children's responses to the predictions of different computational models that can help us better understand what cognitive processes may be involved in multi-modal inference. Given that prior work has found early competence in drawing inferences from sensory information in preschool-aged children (Siegel et al., 2021; Hood, 1995) as well as difficulties in later childhood in multisensory integration (Gori et al., 2008), we targeted a relatively wide age range – from age 3 to 8 – to capture potential developmental change. We expected that while younger participants might understand the task and respond systematically, children may not yet be capable of running mental simulations that integrate vision and sound until early school-aged years.

Models

We consider four different computational models that make different assumptions about the underlying cognitive processes by which children reach their judgment. We illustrate the predictions of each model based on the example shown in Figure 1. The predictions of each model across the nine test trials in the experiment reported below are shown in Figure 2.

Guessing model The simplest possibility is that children might randomly choose one of the three holes. The guessing model implements this prediction by assigning an equal probability to each of the three holes. Such responses may suggest that children did not understand the setup or the task.

Matching model A second possibility is that children might use a matching strategy, and be more likely to choose a hole that is closest in spatial proximity to where the ball landed (ignoring the obstacles in the box). This model is inspired by prior research on the 'gravity bias' which demonstrated children's tendency to assume that objects fall straight down. Children show this bias even when an object is dropped into a curved tube that makes it such that the object ends up in a bucket that's not underneath where it was dropped (Hood, 1995).

The matching model assigns a probability to each hole based on the horizontal distance between the center of the hole and where the ball ended up. The closer the ball is to the center of a hole (along the x-axis), the more probability the model assigns to that hole. The model turns the distance between hole and ball into a likelihood via a Gaussian loss function centered at the location of the ball. We fitted the standard deviation in the loss function to maximize the likelihood of the data. The matching model can be thought of as a graded version of the gravity bias. A child who makes responses that are best explained by this model might have understood the goal of the task, but may have ignored the obstacles in the box as well as the sounds that the ball made. In Figure 1, the matching model predicts that it was most likely that the ball was dropped in hole 3.

Simulation model (vision) Another possibility is that children solve the task by running mental simulations that take into account only the visual information. Gerstenberg, Siegel, & Tenenbaum (2021) developed a computational model of the Plinko task that is illustrated in Figure 1. Their model runs physical simulations of what would happen if the ball

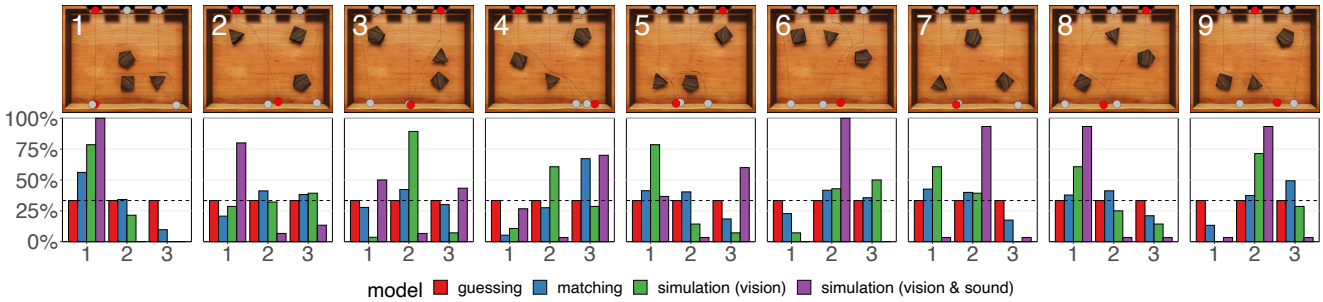


Figure 2: Model predictions across the nine test trials. The boxes at the top show the ball's trajectory for each hole. The red balls indicate where the ball was actually dropped and where it landed. Bar plots under each box show model predictions. The 'guessing' model predicts all holes are equally probable in all trials. The 'matching' model assigns probability for each hole given the horizontal distance between the hole and the ball's final location. The 'simulation (vision)' model infers in which hole the ball was dropped by running simulations that take into account the final position of the ball. The 'simulation (vision & sound)' model also considers the sounds that the ball made when it was dropped. For example, in trial 2, the 'simulation (vision & sound)' believes that the ball was dropped in hole 1, whereas the other models are less certain about what happened.

was dropped into the different holes (Figure 1, middle row) while accounting for the uncertainty in how exactly the ball is dropped ("drop noise") and how it might collide with an obstacle ("collision noise"). In one of their experiments, Gerstenberg, Siegel, & Tenenbaum (2021) tested this model against participants' predictions of where the ball would land if it was dropped in the different holes. They found that participants systematically underestimated how far the ball would go after it collided with an obstacle, and the model captured this by assuming a biased collision noise. The density distributions in the middle row of Figure 1 summarize where the model believes the ball would end up if it was dropped into the different holes. As the density for hole 1 (in blue) shows, the simulation model underestimates how far the ball would go (as indicated by the dotted path that shows the ground truth). In some of the simulated runs, the ball ends up on the left side of the triangle, while in others it ends up on the right side. Based on the simulated trajectories, the model then computes the likelihood of the observed data (i.e. the x-position of where the ball landed in the actual situation) conditioned on each hole by considering how close the ball in each simulated run ended up to where it actually landed. By combining this likelihood with a uniform prior over the different holes, the model computes a posterior belief about where the ball was dropped. In Figure 1, the simulation model with only visual information assigns most probability to hole 3 (bottom left plot).

Simulation model (vision & sound) Finally, we consider the possibility that children are integrating both visual and auditory information in their inferences. Gerstenberg, Siegel, & Tenenbaum's (2021) model encodes the auditory information as a vector that contains the time points at which collisions happened. For example, consider that in the actual situation, the ball collided with an obstacle at $t = 50$, and then landed in the sand at $t = 78$. The model then compares what this vector looks like in each simulated run, with what happened in the actual situation. When the model considers hole 1 in Figure 1, the simulated drops end up generating three sounds (a first collision with the pentagon in the top left, another collision with the triangle, and then the sound of the ball landing in the sand). To compute the likelihood of the auditory data, the model considers how close the sounds of a simulated run match the sounds that were actually heard (the likelihood function includes a penalty when the number of sounds in the simulation doesn't match the number of sounds that actually happened). The model then computes a posterior over the different holes based on a likelihood function that is sensitive to both the visual and auditory evidence. In Figure 1, the simulation model that considers both visual and auditory information infers that the ball must have been dropped in hole 2 (bottom right plot). Even though the ball ended up right underneath hole 3, the fact that there was a collision sound with an obstacle rules out the possibility that it was dropped from that hole.

Experiment

In our experiment, we study children's ability to make inferences from visual and auditory evidence across a number of different Plinko boxes just like the one in Figure 1. The experiment was pre-registered via the OSF (`https://osf.io/rjwqa`). You can access all the materials, data, and analysis here: `https://github.com/cicl-stanford/whats-that-sound`

Methods

Participants We recruited 64 participants between 3 to 8 years of age (Mean_{age} = 5.56; the number of children in each age group ranged from $N = 10$ to $N = 12$) via online advertisements on Facebook and `https://childrenhelpingscience.com`. The demographics of our sample were as follows: *gender*: 32 female, 31 male, 1 preferred not to answer; *race*: 34 White, 11 Asian, 3 Hispanic or Latino, 2 Black or African American, 6 Asian and White, 4

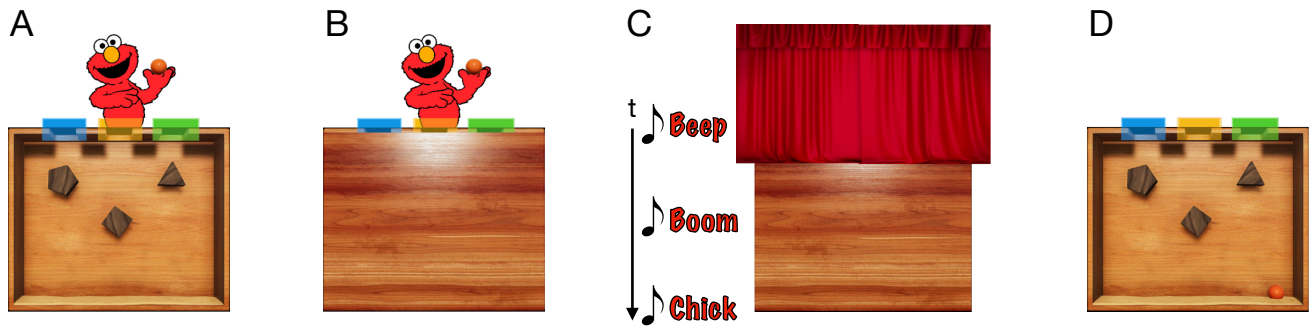


Figure 3: The sequence of events on a test trial. a) Participants saw an open Plinko box (location of obstacles were visible) as Elmo appeared in the middle of the box holding a ball. b) The box was then covered to occlude its contents. c) A curtain was drawn to cover Elmo, and participants heard the sounds of the ball as Elmo dropped it into one of the holes. A 'beep' sound was played when Elmo released the ball. A collision sound ('boom') was played whenever the ball hit an obstacle. A 'chick' sound was played when the ball landed in the sand. In this case, Elmo dropped the ball in hole 2. d) The final location of the ball was revealed and participants were asked: "In which hole did Elmo drop the ball?"

Hispanic/Latino and White, 1 Hispanic/Latino and Pacific Islander, 1 Asian and Hispanic/Latino, 2 participants preferred not to answer. Participants received a $5 gift card as compensation. Twelve additional participants were excluded for failing comprehension check questions (4), environmental interference (1), technical difficulties (4), or opting out (2).

Materials The physical simulations of the Plinko box were created using PyMunk and rendered in 3D with Unity. The sounds were pre-recorded collision sounds that were played at the time at which the collisions happened in the physical simulation. Animations of Elmo interacting with the box were added using Apple Keynote.

Design Participants viewed 9 test trials that were presented in three pseudo-random orders (see Figure 2). The trials varied in the positioning of the obstacles in the box, in which hole Elmo dropped the ball, the number of obstacle collisions (between 0 and 3), and how far the ball lands from the hole in which it was dropped. We selected these trials because the different models make different predictions, thereby allowing us to make inferences about what model is most consistent with individual participants' responses across the trials.

Procedure The experiment was conducted online via Zoom using `https://slides.com/` for stimulus presentation. The study progressed through a familiarization phase, a comprehension check phase, and then a test phase.

In the *familiarization phase*, children first saw Elmo dropping the ball once from each of the three holes in an uncovered box. They saw physically realistic animations of the ball's movements in the box, and also heard the sounds that the ball made as it collided with the obstacles and landed in the sand. This was to ensure that children understood the relevant physical properties of the task. The experimenter then drew children's attention to the different holes (by referring to each hole using the color cues) and the obstacles in the box. The experimenter said that Elmo had many similar boxes that have obstacles in different locations, and that are played with in the same way. Then a second box appeared, with obstacles in different locations, and the experimenter brought children's attention to the sounds. After Elmo dropped the ball in each hole, the experimenter highlighted the different sounds the ball made: 'beep' (when the ball was released), 'boom' (when the ball hit the obstacles or the walls) and 'chick' (when the ball landed in the sand). Each drop was looped three times to ensure that children heard the sounds.

In the *comprehension check phase*, children watched three more animations of Elmo dropping the ball. In each animation, children were able to see in which hole Elmo dropped the ball. After dropping the ball, Elmo disappeared behind the box. In the first animation, the box was uncovered (like in Figure 3a). In the other two animations, the box was covered (like in Figure 3b) so that children could not see the obstacles or the trajectory of the ball as it fell through the box, but could still hear the sounds it made. The cover was then removed so that children saw the final position of the ball (like in Figure 3d). After each animation, children were asked the test question: "In which hole did Elmo drop the ball?". Children responded by saying "blue", "yellow", or "green".

The *test phase* was similar to the comprehension check phase, with one critical difference: a curtain appeared before Elmo dropped the ball, meaning that participants were unable to see in which hole Elmo dropped the ball. Figure 3 illustrates the sequence of events on each test trial.

Results

We will first discuss how children's accuracy in the task develops with age, and then compare children's responses to the predictions of our different models.

Accuracy Figure 4 shows participants' accuracy in the task as a function of age. Overall accuracy increased with age (β = 0.22, 95% credible interval [0.11, 0.33]). However, only 8 year old children's accuracy was reliably above chance (59%

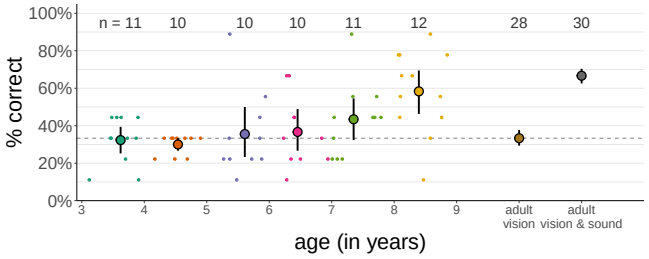


Figure 4: Average accuracy as a function of age. For adults, the results are shown separately for participants who only had visual evidence, or who had both visual and auditory evidence. *Note*: Error bars show 95% bootstrapped confidence intervals for each age group.

[44%, 73%]). It's worth noting though that even adult performance in this task was not at ceiling. As Figure 4 shows, the accuracy of adult participants who only received visual information (i.e. who only saw the final location of the ball but didn't hear the sounds it made when it was dropped) was not above chance. Adult participants who had access to both the visual and auditory information achieved an accuracy of around 70%. Even if children were integrating the visual and auditory information by using mental simulation, we would still expect them to get certain trials wrong. Simulating doesn't necessarily mean getting it right in this task. This is because, as mentioned above, even adult participants tended to underestimate in their simulations, how far the ball would go after it collided with an obstacle. For example, in Figure 2 trial 8, the ball was actually dropped in hole 3. However, the 'simulation (vision & sound)' model assigns the highest probability to hole 1. Because people underestimate how far the ball goes after the collision, hole 1 is a better explanation than hole 3 (and the fact that the sound of the ball colliding with an obstacle happens a little later for hole 1 than for hole 3 is not sufficient to counteract the visual evidence that strongly favors hole 1).

Model comparison To gain more insight into what strategies children were using to solve the task, we compared children's responses to the predictions of our computational models. Figure 5a shows for each participant how well each model captured their responses. To compute the posterior distribution over models, we assumed a uniform prior over the different models, and then computed the likelihood of a participant's responses across the nine trials under each model. Figure 5b shows the posteriors averaged for each age group. Qualitatively, the results show that random guessing was not the predominant response even in the youngest group of children (e.g., only 20% in 3-year-olds); rather, their strategy was more consistent with matching (60%). Additionally, while children's responses were increasingly consistent with the predictions of the mental simulation models, especially between 6 to 8 years of age, the ability to consider both visual and auditory evidence was clear only in 8-year-olds (47%).

Figure 5b also shows adults' inferences across these nine trials. Adults' responses in the vision condition – where they only received visual evidence – were best explained by the 'simulation (vision)' model and the 'matching' model. In the 'vision & sound' condition, where adults heard the sounds that the ball made as it was dropped before they saw its final location, their responses were best explained by the 'simulation (vision & sound)' model.

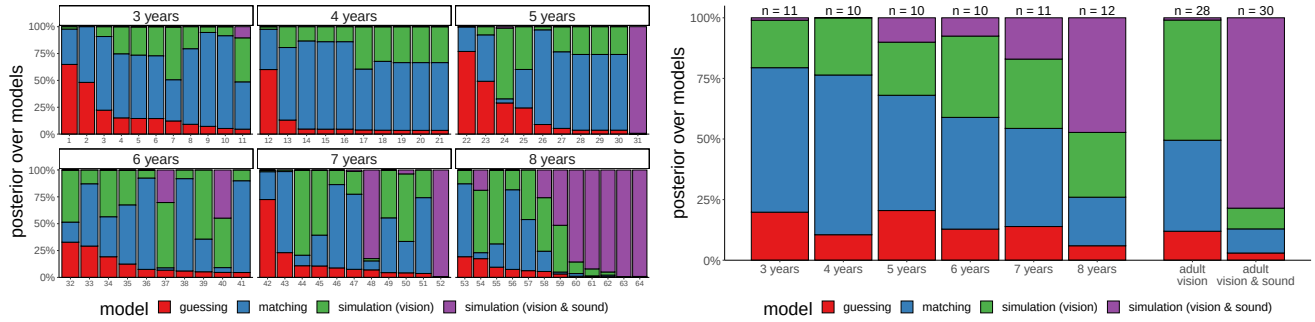
General Discussion

The current work examined whether children can infer what happened by integrating visual and auditory evidence, and how this ability develops in early childhood. We modified the Plinko task (Gerstenberg, Siegel, & Tenenbaum, 2021) to compare 3- to 8-year-olds' behavioral judgments against the predictions of four computational models. The results revealed a clear developmental trend: between preschool and early school-age years, children's accuracy in the task increased with age.

Critically, comparing children's responses against computational models revealed more than just an increase in accuracy. Even though children were performing roughly "at chance" until age 7, our analysis shows that almost none of the children were randomly guessing. Instead, younger children showed a reliable tendency to choose the hole that was closest to where the ball ended up. This matching strategy is consistent with prior work on the gravity bias, suggesting that young children tend to believe that a dropped object ends up straight underneath where it was dropped (Hood, 1995; Tecwyn & Buchsbaum, 2018). There were a number of children (including some 3-year-olds) whose responses were consistent with mental simulation based on visual information. By age 8, the majority of children's responses were consistent with having relied on mental simulation, with a large proportion seemingly having considered both visual and auditory evidence to infer what happened.

In our task, the majority of younger children, before age 6 to 7, relied on a matching strategy, suggesting that they did not reliably engage in simulation to make accurate inferences, and even when they did, failed to integrate vision and sound. This stands in stark contrast to the performance of adult participants who had access to both visual and auditory information, and whose judgments were most consistent with having relied on mental simulation. Given the early-emerging understanding of the physical world (Spelke et al., 1992; Ullman & Tenenbaum, 2020), what makes this task so challenging for children?

First, young children's knowledge of real-world physics may not yet be robust or precise enough to run accurate mental simulations. The Plinko task involves a ball that travels downward, making it particularly challenging for children who may be susceptible to gravity bias. This property of our task may have masked younger children's capacity to run mental simulations. Prior work has shown that the gravity bias disappears when children see videos in which objects



(a) Individual participant model fits

(b) Model fits by age group

Figure 5: Posterior distribution over the different models for a) individual participants (with participant index on the x-axis), and b) across age groups (including model fits for adult participants who either only had visual evidence, or both visual and auditory evidence). Participant 1's responses, for example, were most consistent with 'guessing' (65%), and with 'matching' (33%). This child chose the following holes in the nine trials shown in Figure 2: [1, 2, 2, 3, 2, 2, 1, 1, 3]. Participant 31's responses were most consistent with 'simulation (vision & sound)' (99%). This child chose the following holes: [1, 2, 2, 3, 1, 1, 2, 2, 3]. Across age groups, 3- to 5-year-olds' responses were most consistent with 'matching' (58% on average), whereas 6- to 8-year-olds' responses were increasingly consistent with relying on simulation (55% on average). 8-year-olds' responses were most consistent with having relied on simulation that considers both visual and auditory evidence (47%).

move from the bottom to the top (Hood, 1998), or when children are asked to think about objects moving on a horizontal plane (Hood et al., 2000). Indeed, recent work suggests that even younger children (i.e., 6- to 8-year-olds) may engage in counterfactual simulation to reason about physical events in scenarios that do not involve gravity (Kominsky et al., 2021). A simplified version of our task that minimizes the gravity bias may reveal that even younger children have the ability to infer what happened through mental simulation.

Second, children's ability to integrate visual and auditory information might still be developing during this period. Given prior work showing that 8- to 10-year-olds have difficulty integrating visual and haptic information (Gori et al., 2008), it is not surprising that children in our study had trouble to integrate visual and auditory information before age 7 or 8. While it is difficult to directly compare children's performances across these tasks, the results suggest that multimodal integration may be challenging for younger children. In the Plinko task, visual information is available at the time of judgment but auditory information is not; children in our task had to encode and remember the number, type, and timing of the sounds to figure out what happened. Developing ways to reduce such extraneous demands might be useful for studying children's ability to integrate information from multiple sensory modalities.

Children's responses in our oldest age group (8-year-olds) were more consistent with having relied on a mental simulation that considers both vision and sound than with any of the other models we considered. The extent to which people actually engage in physical reasoning via mental simulation remains an active topic of research. Some argue against this account, suggesting that humans not only employ heuristics to perform complex physical tasks but also exhibit systematic biases in their intuitive physical reasoning (Davis & Marcus, 2016; Ludwin-Peery et al., 2020). Although some effort has been made to accommodate these biases into rational probabilistic Bayesian reasoning models (Zhu et al., 2020), the concerns still persist (Ludwin-Peery et al., 2021). The current work contributes to this debate by exploring the development of these abilities, finding suggestive evidence that children, by 7 to 8 years of age, are capable of engaging in simulation of physical events in ways that integrate multiple channels of sensory information. Additional methods, such as eye-tracking, could shed more light on the role that mental simulation plays in causal inference.

While our model comparison approach provides deeper insights into *how* children may be performing the task, this approach also has some limitations. For example, the current results merely indicate how well different models are doing relative to one another, leaving open the possibility that some children solved the task by relying on a strategy that we didn't consider here. Furthermore, although the four different models represent different strategies for solving the task, the models themselves do not explain what cognitive capacities might develop with age to give rise to this developmental change.

In sum, the human ability to engage in mental simulation has been a longstanding topic of research in cognitive science. The current work suggests that even young children's reasoning about past events may show signatures of simulation that consider not only visual but also auditory information, and that such tendency increases throughout early childhood. Although children's responses may often reflect mistakes and appear to be "at chance", combining computational and developmental approaches can help us look underneath the veil of chance-level performance and provide novel insights into children's ability to reason about what happened and how.

Acknowledgments

Tobias Gerstenberg and Hyowon Gweon were supported by a seed grant from Stanford's Human-Centered Artificial Intelligence Institute (HAI). We thank that Causality in Cognition Lab (CiCL) and the Social Learning Lab (SLL) for valuable feedback on the project. The experiment was approved by Stanford's Institutional Review Board.

References

Agrawal, T., Lee, M., Calcetas, A., Clarke, D., Lin, N., & Schachner, A. (2020). Hearing water temperature: Characterizing the development of nuanced perception of auditory events.

Allen, K. R., Smith, K. A., & Tenenbaum, J. B. (2020). Rapid trial-and-error learning with simulation supports flexible tool use and physical reasoning. *Proceedings of the National Academy of Sciences*, *117*(47), 29302–29310.

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.

Davis, E., & Marcus, G. (2016). The scope and limits of simulation in automated reasoning. *Artificial Intelligence*, *233*, 60–72.

Davis, Z., Allen, K. R., & Gerstenberg, T. (2021). Who went fishing? inferences from social evaluations. In *Proceedings of the 43rd Annual Conference of the Cognitive Science Society.*

Gerstenberg, T., Goodman, N. D., Lagnado, D. A., & Tenenbaum, J. B. (2021). A counterfactual simulation model of causal judgments for physical events. *Psychological Review*, *128*(6), 936–975.

Gerstenberg, T., Siegel, M. H., & Tenenbaum, J. B. (2018). What happened? reconstructing the past from vision and sound. In C. Kalish, M. Rau, J. Zhu, , & T. Rogers (Eds.), *Proceedings of the 40th Annual Conference of the Cognitive Science Society* (p. 409). Cognitive Science Society.

Gerstenberg, T., Siegel, M. H., & Tenenbaum, J. B. (2021). What happened? reconstructing the past from vision and sound. *PsyArXiv*.

Gerstenberg, T., & Tenenbaum, J. B. (2017). Intuitive theories. In M. Waldmannn (Ed.), *Oxford handbook of causal reasoning* (pp. 515–548). Oxford University Press.

Gori, M., Del Viva, M., Sandini, G., & Burr, D. (2008). Young children do not integrate visual and haptic information. *Nature Precedings*, 1–1.

Hood, B. M. (1995, oct). Gravity rules for 2- to 4-year olds? *Cognitive Development*, *10*(4), 577–598.

Hood, B. M. (1998). Gravity does rule for falling events. *Developmental Science*, *1*(1), 59–63.

Hood, B. M., Santos, L., & Fieselman, S. (2000). Two-year-olds' naïve predictions for horizontal trajectories. *Developmental Science*, *3*(3), 328–332.

Kim, M., & Schachner, A. (2021). From music to animacy: Causal reasoning links animate agents with musical sounds.

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.

Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. *Behavioral and Brain Sciences*, *40*, 1–72.

Lopez-Brau, M., Kwon, J., & Jara-Ettinger, J. (2020). Mental state inference from indirect evidence through bayesian event reconstruction. In *Proceedings of the 42nd Annual Conference of the Cognitive Science Society* (pp. 467–473).

Ludwin-Peery, E., Bramley, N. R., Davis, E., & Gureckis, T. M. (2020). Broken physics: A conjunction-fallacy effect in intuitive physical reasoning. *Psychological Science*, 0956797620957610.

Ludwin-Peery, E., Bramley, N. R., Davis, E., & Gureckis, T. M. (2021). Limits on simulation approaches in intuitive physics. *Cognitive Psychology*, *127*, 101396.

Schachner, A., & Kim, M. (2018). Alternative causal explanations for order break the link between order and agents.

Siegel, M. H., Magid, R. W., Pelz, M., Tenenbaum, J. B., & Schulz, L. E. (2021). Children's exploratory play tracks the discriminability of hypotheses. *Nature Communications*, *12*(1).

Smith, K. A., & Vul, E. (2012). Sources of uncertainty in intuitive physics. In *Proceedings of the 34th Annual Conference of the Cognitive Science Society.*

Smith, K. A., & Vul, E. (2014). Looking forwards and backwards: Similarities and differences in prediction and retrodiction. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th Annual Conference of the Cognitive Science Society* (pp. 1467–1472). Austin, TX: Cognitive Science Society.

Spelke, E. S., Breinlinger, K., Macomber, J., & Jacobson, K. (1992). Origins of knowledge. *Psychological Review*, *99*(4), 605–632.

Tecwyn, E. C., & Buchsbaum, D. (2018). Hood's gravity rules. In J. Vonk & T. Shackelford (Eds.), *Encyclopedia of animal cognition and behavior.* Springer, Cham.

Ullman, T. D., Spelke, E., Battaglia, P., & Tenenbaum, J. B. (2017). Mind games: Game engines as an architecture for intuitive physics. *Trends in Cognitive Sciences*, *21*(9), 649–665.

Ullman, T. D., & Tenenbaum, J. B. (2020). Bayesian models of conceptual development: Learning as building models of the world. *Annual Review of Developmental Psychology*, *2*, 533–558.

Wu, Y., Schulz, L. E., Frank, M. C., & Gweon, H. (2021). Emotion as information in early social learning. *Current Directions in Psychological Science*, *30*(6), 468–475.

Zhu, J.-Q., Sanborn, A. N., & Chater, N. (2020). The bayesian sampler: Generic bayesian inference causes incoherence in human probability judgments. *Psychological review*, *127*(5), 719.