The trajectory of counterfactual simulation in development

Jonathan F. Kominsky[1], Tobias Gerstenberg[2], Madeline Pelz[3], Mark Sheskin[4], Henrik Singmann[5], Laura Schulz[3], & Frank C. Keil[4]

[1]Harvard University, [2]Stanford University, [3]MIT, [4]Yale University, [5]University of Warwick, UK

Abstract

Previous work has argued that young children do not answer counterfactual questions (e.g. "what would have happened?") by constructing simulations of alternative possibilities in the way adults do. Here, we propose that children can engage in simulation when answering these questions, but consider different counterfactual possibilities than adults. While most previous research has relied on narrative stimuli, we use causal perception events, which are understood even in infancy. In Experiment 1, we replicate earlier findings that children struggle with counterfactual reasoning, but show that they are capable of conducting the required simulations in a prediction task. In Experiment 2, we use a novel multiple-choice method that allows us to study not only *when* children get it right, but also *how* they get it wrong. We find evidence that 4-year-olds engage in simulation, but preserve only some features of what actually happened and not others.

Keywords: causality; counterfactual reasoning; perception; child development; multinomial process trees

Introduction

When considering whether some event C caused another event E, we do not merely consider events as they actually unfolded. Rather, we think about what *could* or *would* have happened had C been altered in some way (Byrne, 2016; Lewis, 1973). This capability for counterfactual reasoning is an essential, and perhaps even automatic, feature of causal cognition (Gerstenberg, Peterson, Goodman, Lagnado, & Tennenbaum, 2017), with a variety of consequences. For example, the relevance of different counterfactual possibilities affects causal judgments (Phillips, Luguri, & Knobe, 2015; Icard, Kominsky, & Knobe, 2017; Phillips & Kominsky, 2017), counterfactual reasoning undergirds emotions, like regret and relief (Beck & Riggs, 2014), and is an implicit component of Bayesian causal learning (Pearl, 2000).

One of the essential properties of counterfactual reasoning is *simulation*. When people engage in counterfactual reasoning, they construct a mental model of the events as they actually happened, and then imagine how the events would have unfolded if something about the situation had been different. This mental simulation is guided by a causal model of the situation which dictates the consequences of counterfactual interventions (e.g., Sloman & Lagnado, 2005).

The developmental origins of counterfactual reasoning in the human mind remain a challenging mystery to cognitive science. Piaget held that counterfactual reasoning emerged in the developmental stage of "formal operations", starting at about 12 years of age (Inhelder & Piaget, 1958). Later work found that children as young as 3 could answer certain counterfactual questions correctly. For example, presented with a story about a girl named Carol who walked across a floor with dirty shoes, 3-5-year-olds who were asked "what would have happened if Carol had taken her shoes off?" correctly answered the floor would be clean (Harris, German, & Mills, 1996).

However, later work suggested that children may arrive at such answers without engaging in counterfactual simulation, and simply rely on conditional reasoning instead. In general, dirty shoes make floors dirty, while clean shoes leave floors clean (Rafetseder, Schwitalla, & Perner, 2013). However, basic conditional reasoning and counterfactual reasoning come apart in situations in which the outcome is causally *overdetermined*. When an outcome was overdetermined, this means that there were multiple individually sufficient causes such that the outcome would still have come about even if one (or more) of the causes hadn't occurred. For example, if both Carol and Max walk across the kitchen floor with dirty shoes, and children and adults are asked what would have happened if Carol had taken her shoes off, adults say the floor would still have been dirty (because of Max), whereas 5-year-olds overwhelmingly say the floor would have been clean. Remarkably, 10-year-olds responded at chance, and adult-like performance emerged only around 14 years of age (Rafetseder et al. 2013).

Recent work has, again, been more optimistic about children's counterfactual reasoning abilities. When narratives are replaced by simple "blicket detector" causal systems in which only some blocks (called "blickets") can make a machine go, children show above-chance success for overdetermined outcomes around age 6 (Gopnik & Sobel, 2000), or even at age 4-5 (Nyhout & Ganea, 2019).

However, we believe that what it means to succeed in counterfactual reasoning needs to be examined more closely. In the research to date, researchers have generally concluded that the reason why children answer these questions incorrectly, is because they *do not simulate* counterfactual alternatives, but instead arrive at their answers by some other reasoning strategy (Rafeteseder et al., 2013; Nyhout & Ganea, 2019). This is remarkable given that other work has found that children are quite adept at simulation when making *predictions* about events that have not yet occurred (Atance & O'Neill, 2005). Given that children can engage in simulation in some cases, and that adults naturally do so when answering causal questions (Gerstenberg et al., 2017), the assumption that young children fail to reason counterfactually because they do not engage in counterfactual simulation *at all* is worth re-examining.

There is another possible reason for why children respond differently than adults: Rather than failing to simulate, they instead simulate different counterfactual alternatives than adults do. This proposal aligns with a recent proposal that young children may consider a broader hypothesis space than adults do when engaging in causal reasoning (Gopnik et al., 2017). Similar to how children may be more flexible in what hypotheses they consider in causal reasoning, it is possible that they also consider different possibilities than adults do, when simulating counterfactuals. Here, we are interested to see whether there is systematicity in the way in which children consider counterfactual possibilities. When children get the answer to a counterfactual questions wrong, are they just randomly guessing, or may they systematically consider different possibilities than adults do? Characterizing such potential systematicity could give unique insight into the development of counterfactual reasoning, and a deeper understanding of what features of an event children consider *mutable* (Byrne, 2016).

In order to examine which specific counterfactual possibilities children consider, we depart from the narrative studies that have been used in most prior work. Narrative stimuli add a great deal of memory load and room for influence from idiosyncratic knowledge. The ideal stimuli would be a causal event that children understand nearly effortlessly, that they can see in full while answering a counterfactual question, and which offers the opportunity to ask not just whether they are simulating counterfactual alternatives, but which specific alternatives they consider.

Simple physical interactions that fall under the category of "causal perception" perfectly fit these criteria. Events in which one object appears to collide with another and cause it to move are perceived as causal by 6 *months* of age (Leslie & Keeble, 1987; Saxe & Carey 2006; Kominsky et al., 2017), and recent work has used these events to demonstrate counterfactual simulation in causal judgment with adults (Gerstenberg et al., 2017).

In the current work, we present two experiments investigating the development of counterfactual simulation, using causal perception events. In Experiments 1a and 1b, we replicate previous findings that children struggle with counterfactual reasoning in overdetermined cases, but in the domain of causal perception events. However, we also find that children are highly accurate when making *predictions* about these events, showing that they are able, in principle, to conduct the necessary simulations to answer the questions correctly.

In Experiment 2, we present children with concrete counterfactual alternatives to causal perception events in a multiple-choice answer format similar to that employed by Rafetseder and Perner (2018). This response format allows us to examine not only *whether* children engage in counterfactual simulation, but *which specific counterfactual possibilities* they consider.

Experiment 1a

The goal of this experiment was to validate the domain of causal perception in the study of children's counterfactual judgments, by having children make counterfactual judgments about simple causal perception events.

Methods

Participants We planned to run 40 children in each age group (20 in each of two conditions), and continued collecting data until we had reached that target, replacing any participants that were excluded (see below). 40 5-6-year-olds (15 female), 40 7-8-year-olds (15 female), and 40 9-10-year-olds (18 female) participated in Experiment 1a, recruited from local schools and children's museums. In addition, 10 5-6-year-olds (5 female), 3 7-8-year-olds (2 female) and 1 (male) 9-10-year- old participated but were excluded from analyses based on predetermined exclusion criteria (see below).

Stimuli and procedure We constructed simple animations modeled on those used by Gerstenberg et al. (2015) (see Fig. 1, videos of the animations can be found here: http://osf.io/qwphr/). In these animations, there are two balls, A and E, a red area that was described as a "goal", and black walls on either side of the goal. The stimuli were animated .gif files placed into a Qualtrics survey (Qualtrics, 2005). The survey was presented on an iPad.

All participants first saw two training items in counterbalanced order. In one training item, ball A hit ball E, which then bounced off the wall above the goal. In the other training item, ball A hit ball E, which then went into the goal. Following each training trial, participants were asked two questions: "Before ball A hit ball E, was ball E moving or sitting still?", and "Did ball E go into the goal?" Participants could verbally respond and the experimenter would record their answer, or older children could select the option on the iPad directly. If participants answered either question incorrectly on one of the training trials, they were shown that training animation a second time and asked again.

Participants then saw one of two test trials, between-subjects. In the "difference-making" condition, the animation was almost identical to the training item in which ball E bounced off the wall above the goal, except that there was a "brick wall" (see Fig. 1) that ball E bounced off of, and ball E went into the goal. In the "overdetermined" condition, the animation was almost identical to the training item in which the ball went into the goal, except that the ball bounced off the brick wall before going into the goal, thus leaving the outcome unchanged.

Following the test trial, participants were asked the same two questions as in the training trials. If children answered either question incorrectly, they were not corrected but their data were excluded. Then, children were asked the critical test question: "What if the brick wall had not been there? Would ball E have gone into the goal?"

Results and discussion

Results can be found in Fig. 2. A simple inspection of this figure gives a clear sense of the results, which were similar across all age groups: Near-perfect performance on cases in which the brick wall made a difference (where the correct answer is that ball E would not have gone into the goal), but only roughly 50% accuracy for overdetermined events (where the correct answer is that ball E would still have gone

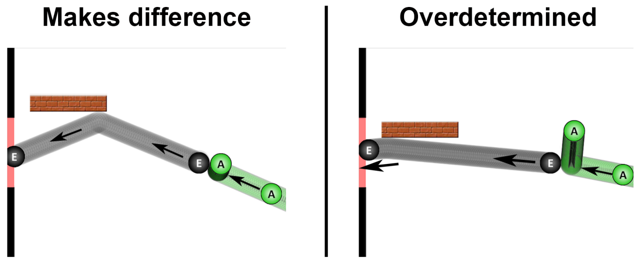

Figure 1. Example stimuli from Experiment 1a. In the difference-making event (left), the brick wall altered ball E's trajectory such that it went into the goal. In the overdetermined condition (right), ball E also deflects of the wall, but would have gone into the goal regardless.

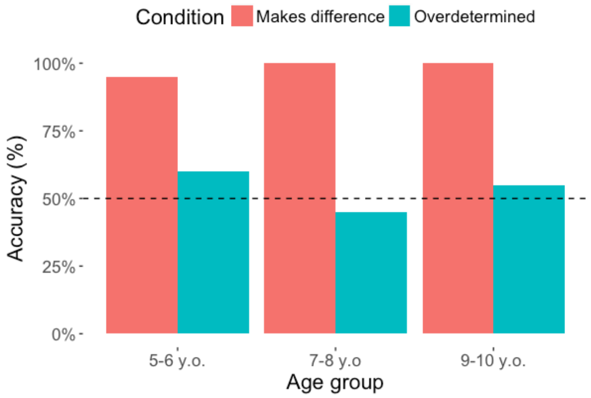

Figure 2. Proportion of accurate responses to the counterfactual question in Experiment 1a.

into the goal). A logistic regression with age group and condition as factors revealed a main effect of condition, ß = 2.54, $p = .02$, but no effect of age group and no interactions, $p > .9$. As children demonstrated nearly uniform perfect performance in the difference-making condition (one incorrect answer in total), no further analyses were conducted for this condition. For the overdetermined condition, a logistic regression with age group also showed no effect of age ($p > .3$) and no significant intercept (p = .37), indicating that accuracy did not differ from chance (i.e., .5).

These results are very similar to many earlier results investigating children's counterfactual reasoning (e.g., Rafetseder et al., 2013): Children can answer counterfactual questions when the correct answer changes the outcome, but struggle in overdetermined cases. One reason for this could be that children are unable to successfully simulate the required counterfactual possibility in these causal perception events. Experiment 1b tested this hypothesis by asking children to make predictive simulations about these very events, without the brick wall.

Experiment 1b

In this experiment, we wanted to see whether children are capable of correctly predicting what will happen after the animation is paused. It is possible that children failed to answer the counterfactual question correctly in the overdetermined situation because they have trouble simulating what would have happened in this case.

Methods

Participants This study was stopped early due to the fact that all children responded correctly. Our final sample sizes were therefore 21 5-6-year-olds (10 female) and 26 7-8-year-olds (14 female) recruited from the same populations as Experiment 1a. In addition, 4 5-6-year-olds (2 female) and 1 (male) 7-8-year-old were excluded based on predetermined exclusion criteria (see below).

Stimuli and procedure The stimuli were similar to Experiment 1a with the following differences: Participants first saw four training trials in random order: Two in which ball E went into the goal and two in which it missed the goal. First, children saw an animation where ball A struck ball E, and ball E moved approximately halfway from its starting position to the left edge of the display (where the wall and goal are located). At this point the animation froze and a large "pause" icon appeared (that didn't obstruct either of the balls). Children were then asked, "If ball E keeps going, will it go into the goal?" Children could respond "yes" or "no". For the training trials, children then saw the rest of the animation. If children made incorrect predictions on at least two of these items, they were excluded from analyses on the basis that they did not understand the task.

Following training, children saw two test trials, a "difference-making" trial and an "overdetermined" trial in counterbalanced order. The test trials were identical to those used in Experiment 1a, with two exceptions: First, the brick wall was not visible (i.e., identical to Experiment 1a's training trials). Second, the animation paused on the frame in which the ball would have collided with the brick wall in Experiment 1a (participants had no way of knowing this). Participants were then asked the same question as in the training items, but were not shown the end of the animation. Note that the predictions that children are asked to make in Experiment 1b are identical to the counterfactual simulation that is required to answer what would have happened without the brick wall in Experiment 1a.

Results and discussion

Every single child who passed the training provided correct answers to both test questions (21/21 5-6-year-olds and 26/26 7-8-year-olds). We report no statistical tests because the uniformity of these responses renders such tests uninformative.

Experiment 2

Experiment 1b showed that, in line with prior work, children are capable of engaging in the kind of physical simulation that is required to answer counterfactual questions correctly, but did not do so consistently for the overdetermined item in Experiment 1a. This result suggests that children's counterfactual reasoning about causal perception stimuli is similar to their reasoning in other domains. However, we cannot tell based on these findings why children sometimes get it wrong. One explanation is like that proposed by Rafetseder et al. (2013): Children did not engage in simulation at all when asked to consider the counterfactual question. While this is still possible, given that they are obviously capable of engaging in simulation, we must ask *why*.

One possibility is that children cannot simulate while holding the event as it actually occurred in mind (Beck & Riggs, 2014). For example, a correct answer in Experiment 1a requires mentally rewinding the animation and then simulating what would have happened without the brick. The corresponding prediction in Experiment 1b is simpler because the brick is not present in the scene, the clip is paused, and it only requires children to simulate the future without the need to go back in time.

An alternative is that the wording of the question influenced children's performance. Notably, we found a pattern that aligns more closely with Rafetseder et al. (2013) than more recent work (Nyhout & Ganea, 2019; Rafetseder & Perner, 2018). One key difference between our study and that of Nyhout and Ganea (2019) is that the question in Nyhout and Ganea was "would [outcome] *still* [have happened]?" (emphasis added). While a systematic investigation is necessary, children may sometimes be answering on the basis of pragmatic cues: Why ask "would the outcome have been different" if the outcome was unchanged?

A second, not mutually exclusive alternative is that children *did* engage in simulation, but considered different counterfactual possibilities than adults did. We explore this possibility in Experiment 2 using a multiple-choice task modeled on Rafetseder and Perner (2018). We hypothesized that children may arrive at the wrong answer because they hold some of the features of the actual event constant, but allow other features to vary in ways that adults and older children do not. Based on pilot data, we predicted that children will specifically maintain the point of origin of a ball's movement from the event they saw, much as we would expect adults to, but allow the initial trajectory of the ball to vary, which we would not expect adults to do.

Methods

Participants We pre-registered (https://osf.io/qn3b9) a planned sample size of 24 participants in each of three age groups: 4-year-olds, 5-year-olds, and 6-year-olds. We therefore recruited 24 4-year-olds (15 female), 24 5-year-olds (7 female) and 24 6-year-olds (8 female). In addition, 6 4-year-olds (2 female) and 2 (female) 5-year-olds participated but were excluded due to failing to complete the study (4) or parental interference (3; see below). Participants were recruited from TheChildLab.com (Sheskin & Keil, 2018).

Stimuli and apparatus Children saw a total of ten trials in which featured animated events, and then still images representing what actually occurred in the animation, as well as four counterfactual possibilities (see Fig. 3; Full stimuli are available online at https://osf.io/5jw6y/).

Animated events were constructed using Flash, converted to a movie format, embedded in a PowerPoint presentation, and presented over a videoconferencing system. The animations were slightly modified from Experiment 1. This time, there was only one ball, resembling a soccer ball, and the brick wall was replaced with a triangular wedge with a wood texture. The background was made green with a white line to mimic a soccer field. The goal was turned into a grey rectangle, and there were no walls on either side of it.

We created a total of eight test animations and two training animations. In all test animations, the ball entered the stage from the right side and moved in a perfectly horizontal trajectory. In six of the test animations, the ball deflected off of the wedge, which did (4 animations) or did not (2) change whether it went into the goal. In two other test animations, the ball did not interact with the wedge, and simply moved across the field in a straight line.

Along with each test animation, we made a still image that showed the entire trajectory the ball had taken (center, Fig. 3), which was visible while the child was answering the counterfactual question, thus eliminating memory load. In addition, we constructed still images representing four counterfactual possibilities for each animation (Fig. 3). In these counterfactual possibilities, the wedge was removed, and the complete trajectory of the ball was shown as in the still image of the actual event. These four possibilities were constructed in systematic ways for the six items in which the ball interacted with the wedge.

- "**Correct**" (red): In this image, the ball is shown moving horizontally across the entire field, starting from the same point of origin that it had in the actual animation. In other words, it preserved both the origin and the initial trajectory of the ball.
- "**Match origin**" (yellow): The ball started from the same point of origin, but had a diagonal trajectory, ultimately ending up in the exact same place as the ball ended up in the actual event, in which it deflected off the wedge. This option preserved the origin but not the trajectory of the actual event.
- "**Match trajectory**" (purple): The ball originated from a y-coordinate that was level with where the ball *ended* in the actual animation, and the ball moved across the whole field in a perfectly horizontal trajectory. This option preserved the initial trajectory but not the origin of the actual event.
- "**Match neither**" (blue): The ball started from the same place as it did in the "match trajectory" image, but had a diagonal trajectory ending in the same place as the "correct" image, thus matching neither the point of origin nor the initial trajectory of the actual event.

For the events in which the ball and wedge did not interact, the four images still contained two options that preserved the origin and two that preserved the trajectory, but because the

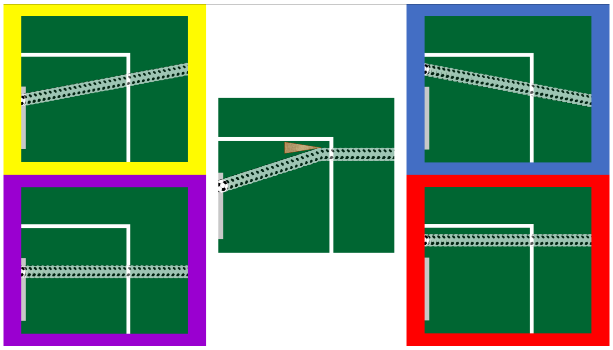

Figure 3. Example item from Exp. 2, as a child would see it. The center image is a rendering of the video the child just watched. On this trial, red is "correct", yellow is "match origin", purple is "match trajectory", and blue is "match neither".

ball did not deflect off the wedge in the actual event, the "match origin" and "match trajectory" images in fact showed the ball ending up in a location that was not present in the original event, while the "correct" and "match neither" images did. The model we used to analyze children's responses (described below) therefore does not apply to these images.

In addition, there were two training animations, one in which the ball bounced off the wedge and one in which it did not interact with the wedge. In both training animations, the ball entered on a diagonal trajectory. No still image of the event was presented in the center of the response screen, and in the still images for training items, the wedge was still present, as the training task was matching the *actual* event rather than considering a counterfactual one.

Procedure The script can be found in the presenter notes of the PowerPoint presentations at http://osf.io/5jw6y/

After parents gave informed consent, children were first shown the two training animations, and after each one asked to find the image that matched what they saw from the four possibilities. This was primarily to familiarize children with the multiple-choice response method. For test trials, children were asked "If there were no block on the field, how would the ball have moved?"

The experimenter was blind to what the child was seeing at all times, and only recorded the color that they said. Children's responses were then transcribed by another coder who was blind to condition, and later matched to images based on the condition the child had been assigned to (see data files in repository). There were two exclusion criteria: If the child failed to finish the study for any reason, or if the parent interfered in a way that guided the child toward a specific answer on any item, in the opinion of the experimenter or coder. As both were blind to what the child was seeing, these judgments could not be influenced by knowing what option the child was selecting.

Analysis plan We focused on the six test items in which the ball collides with the wedge. For those items, we used a multinomial processing tree (MPT) model (Riefer & Batchelder, 1988) to model the proportions with which the different age group chose the four possible response options, *P*(*C*) (= "Correct"), *P*(*O*) (= "match origin"), *P*(*T*) (= "match trajectory"), and *P*(*N*) (= "match neither"). Our model has three free parameters, $\boldsymbol{s}$, $\boldsymbol{m_o}$, and $\boldsymbol{m_t}$, which each represent the (conditional) probability of reaching a specific discrete cognitive processing stage (e.g., $\boldsymbol{s}$ = probability of engaging in simulation). In addition, our model allowed for the possibility of unbiased guessing.

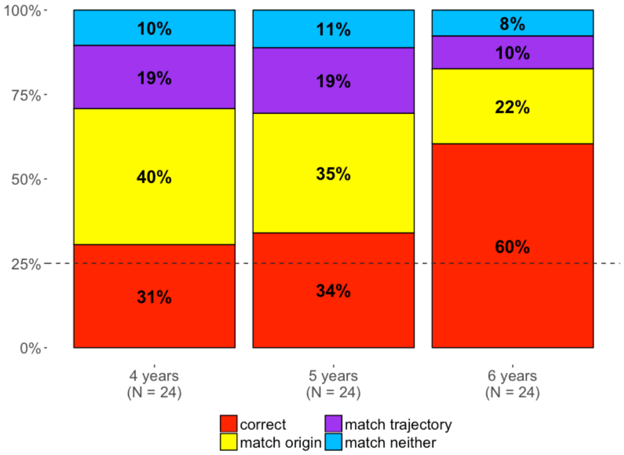

Figure 4. Results of Exp. 2. Proportion of responses is on the y-axis, and chance responding is indicated at 25%.

The first parameter (*s*) represents the probability whether the children engage in simulation or not. If they do not (with probability 1–*s*), we assume children simply make an unbiased guess for one of the four response categories (i.e., the conditional probability of choosing any one response category is .25). We assume that this will be unbiased as the multiple-choice question lacks the pragmatic demands of the questions used in previous studies. In case children engage in simulation (with probability *s*), we assume two further (unordered) processing steps: how likely they are to maintain the origin from the actual world in their simulation (parameter m_o), and how likely they are to maintain the trajectory (parameter m_t)? In order to examine this, we ignore the cases in which the ball does not interact with the block. For the remaining six cases, we can enumerate how the four different response categories follow from the assumed processes. For example, if children maintain both the origin and the trajectory (with probability $m_o \times m_t$), they will provide the correct response. If, however, children only maintain the origin, but not the trajectory (with probability $m_o \times (1 - m_t)$), they will choose the "match origin" response option, $P(O)$. Analogous arguments can be made for $P(T)$ and $P(N)$. Thus, the following model equations are assumed to hold:

$$P(C) = s \times m_o \times m_t + (1 - s) \times 0.25$$
$$P(O) = s \times m_o \times (1 - m_t) + (1 - s) \times 0.25$$
$$P(T) = s \times (1 - m_o) \times m_t + (1 - s) \times 0.25$$
$$P(N) = s \times (1 - m_o) \times (1 - m_t) + (1 - s) \times 0.25$$

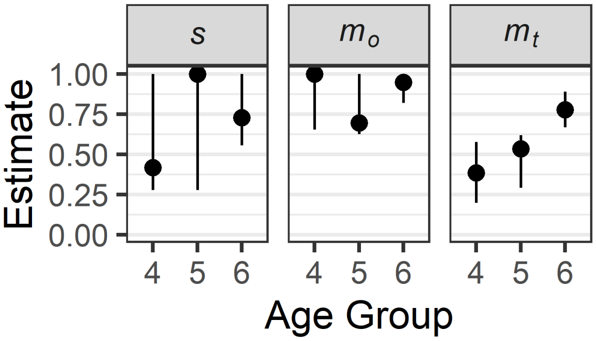

Figure 5. MPT model parameter estimates for s, m_o, and m_t in each age group. Error bars are bootstrapped 95% CIs.

To obtain estimates of the model parameters s, m_o, and m_t, we fitted the model to the aggregated data using maximum-likelihood estimation. This provides us with a model-based estimate of how likely each age group is engaging in simulation, and the likelihood, in each age group, of maintaining the origin and the trajectory of the actual world. Note that, although the model is saturated (i.e., three free parameters for three independent data points provided by the multinomial distribution with four categories), it cannot account for any possible data pattern. That is, our model imposes testable constraints on the data. For example, it predicts that, after accounting for the proportion of unbiased guessing, the conditional ratio of $P(C)/P(O)$ must be equal to the conditional ratio $P(T)/P(N)$. Therefore, finding that the model adequately accounts for the data (i.e., it fits the data), provides some evidence for the underlying assumptions and validity of the interpretations associated with the parameters.

Results and discussion

Fig. 4 shows how often children chose each of the four options for the six test items where the ball collided with the wedge. For the two cases in which the ball and wedge did not interact, the correct answer was the modal response in every age group (4-year-olds: 50%; 5-year-olds: 71%; 6-year-olds: 88%).

A visual inspection of the figure suggests a clear pattern when it comes to choosing the correct answer: Above-chance performance emerges around age 6. However, it also appears that, of the three possible incorrect responses, all age groups preferred “match origin” over “match trajectory” and “match neither”, which suggests that the younger children are not just guessing randomly. Rather, they are simulating possibilities that maintain the origin but not the trajectory of the ball in the actual event.

To verify this impression, we fit our MPT model to children’s responses. As our model was saturated, we used a double bootstrap procedure (van de Schoot, Hoijtink, & Dekovic, 2010) to evaluate model fit. This approach revealed a p-value of .04 ($G^2 = 3.48$) for the 4-year olds and .05 ($G^2 = 2.66$) for the 5-year-olds, suggesting that the main patterns in the data were well accounted for, but there was some misfit. Specifically, the model cannot predict both $P(C) < P(O)$ and $P(T) > P(N)$ at the same time, as was observed in the data. One possible reason for this misfit is individual differences in the simulation behavior of the 4- and 5-year-olds, such that some individual children consistently responded in a particular way and others did not. For 6-year-olds, the fit was perfect ($G^2 = 0$). Given the small magnitude of misfit, the model is interpretable, and we can evaluate the likelihood that children engaged in simulation, and how.

The parameter estimates for each parameter in each age group, with 95% confidence intervals estimated by parametric bootstrapping, can be seen in Fig. 5. In short, we find little evidence for developmental change in m_o or s, but a clear developmental increase in the estimate of m_t. Put in plain terms, this analysis suggests that 4- and 5-year-olds were not significantly different from 6-year-olds or each other in their likelihood of engaging in simulation, nor in how likely they were to choose an option that maintained the ball’s point of origin from the actual event. However, 6-year-olds were significantly more likely than younger children to maintain the ball’s initial *trajectory* from the counterfactual event. In addition, for 6-year-olds we have considerably smaller CIs for s, indicating we that we have higher certainty that they engage in simulation most of the time.

In short, children ages 4-5 do seem to engage in counterfactual simulation, and systematically hold constant some, but not all, features of the actual world in those counterfactual simulations, while allowing other features of the world (which older children hold constant) to vary.

General Discussion

In two experiments, we provide evidence that young children engage in counterfactual simulation, but do so in a different way than older children and adults. Experiment 1 validated the stimuli by replicating previous findings about children’s ability to answer counterfactual questions and conduct predictive simulations, but in the domain of causal perception. Experiment 2 asked children to choose among four counterfactual trajectories rather than answering a simple yes/no question, and found that when 4-5-year-old children engage in simulation, they consider counterfactual possibilities in which the origin of an object’s motion is preserved while its initial trajectory is allowed to vary, while 6-year-olds are more likely to preserve both features in their counterfactual simulations.

We consider these findings in the context of a general theory of children’s reasoning put forward by Gopnik et al. (2017): When reasoning about different possibilities, children’s hypothesis space may be quite different from adults, but the basic process of simulation could be very similar. In particular, this theory suggests that children have a broader and “flatter” hypothesis space (i.e., priors across all hypotheses are similar), in which they conduct a “higher-temperature” (i.e., broader) search. This theory can be readily applied to children’s struggles with counterfactual reasoning: When considering counterfactual possibilities, children may be sampling from a broader set of possibilities, none of which are favored over the others, and the way they pick possibilities out of this space is more random.

However, unlike Gopnik et al. (2017), we find that children are only showing evidence of this broader search space for certain specific features of these events. In other words, while our results align with the general proposal that children conduct simulations over a “flatter” hypothesis space, the space is only flatter over certain “dimensions” (i.e., components) of the events being considered. In this case, children are unlikely to consider possibilities that change where the ball enters from, but between 4 and 6 they narrow the search space for the initial trajectory of the object to be more like adults’.

This is a critical advance for understanding children’s reasoning. We must not only test whether they are searching a broader space of possibilities in general, but also identify the separate features of that hypothesis space and determine which aspects of the event-structure are treated in an adult-like way (in this case the point of origin of the object’s motion). Doing so will not only help us better understand children’s reasoning processes, but allow us to predict specific challenges they face, or errors they will make.

One limitation is that we selected the range of possibilities for children to consider, and so there may be a possibility that we did not include which they would prefer over and above the ones they selected here. While verbal pragmatics are no longer a viable explanation, there are other possible explanations for children’s responses that would not rely on simulation, such as path similarity, or some kind of contextual inference about the scenario, such as whether there is an agent launching the ball into motion.

This work provides an exciting new approach to the study of counterfactual reasoning in development. We should consider that “failure” in these tasks may result not from a failure to simulate *per se* but rather from different assumptions about what to hold constant and what to change when simulating counterfactuals.

References

Atance, C. M., & O’Neill, D. K. (2005). The emergence of episodic future thinking in humans. *Learning and Motivation*, *36*(2), 126-144.

Beck, S. R., & Riggs, K. J. (2014). Developing Thoughts About What Might Have Been. *Child Development Perspectives*, *8*(3), 175-179. doi:10.1111/cdep.12082

Byrne, R. M. J. (2016). Counterfactual Thought. *Annual Review of Psychology*, *67*, 135-157.

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

Gerstenberg, T., Goodman, N. D., Lagnado, D. A., & Tenenbaum, J. B. (2015). How, whether, why: Causal judgments as counterfactual contrasts. In *Proceedings of the 37th Annual Conference of the Cognitive Science Society*, Austin, TX, 2015 (pp. 782--787). Cognitive Science Society.

Gopnik, A., O’Grady, S., Lucas, C. G., Griffiths, T. L., Wente, A., Bridgers, S., . . . Dahl, R. E. (2017). Changes in cognitive flexibility and hypothesis search across human life history from childhood to adolescence to adulthood. *Proceedings of the National Academy of Sciences*, *114*(30), 7892-7899.

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

Icard, T. F., Kominsky, J. F., & Knobe, J. (2017). Normality and actual causal strength. *Cognition*, *161*, 80-93.

Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking: From childhood to adolescence.* (A. Parsons, S. Milgram, Trans.). New York, NY: Basic Books.

Kominsky, J. F., Strickland, B., Wertz, A. E., Elsner, C., Wynn, K., & Keil, F. C. (2017). Categories and Constraints in Causal Perception. *Psychological Science*, *28*(11), 1649-1662. doi:10.1177/0956797617719930

Leslie, A. M., & Keeble, S. (1987). Do six-month-old infants perceive causality? *Cognition*, *25*(3), 265-288.

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

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

Pearl, J. (2000). *Causality : models, reasoning, and inference*. Cambridge, U.K.; New York: Cambridge University Press.

Phillips, J., Luguri, J. B., & Knobe, J. (2015). Unifying morality’s influence on non-moral judgments: The relevance of alternative possibilities. *Cognition*, *145*, 30-42.

Phillips, J., & Kominsky, J. F. (2017). *Causation and norms of proper functioning: Counterfactuals are (still) relevant.* Proceedings from Proceedings of the 39th annual meeting of the cognitive science society.

Qualtrics. (2005). [Computer Software]. Provo, UT: Qualtrics.

Rafetseder, E., & Perner, J. (2018). Belief and Counterfactuality. *Zeitschrift für Psychologie*, *226*, 110-121.

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

Riefer, D. M., & Batchelder, W. H. (1988). Multinomial modeling and the measurement of cognitive processes. *Psychological Review*, *95*(3), 318–339.

Saxe, R., & Carey, S. (2006). The perception of causality in infancy. *Acta Psychologica*, *123*(1-2), 144-165.

Sheskin, M., & Keil, F. (2018). TheChildLab.com A Video Chat Platform for Developmental Research. Retrieved from psyarxiv.com/rn7w5

van de Schoot, R., Hoijtink, H., & Dekovic, M. (2010). Testing Inequality Constrained Hypotheses in SEM Models. *Structural Equation Modeling: A Multidisciplinary Journal, 17*(3), 443–463.