CHILD DEVELOPMENT

Child Development, xxxx 2021, Volume 00, Number 0, Pages 1–16

Thinking Counterfactually Supports Children s Evidence Evaluation in Causal Learning

Jae Engle
University of California, San Diego and Scripps Research

Caren M. Walker
University of California, San Diego

Often, the evidence we observe is consistent with more than one explanation. How do learners discriminate among candidate causes? The current studies examine whether counterfactuals help 5-year olds ($N = 120$) select between competing hypotheses and compares the effectiveness of these prompts to a related scaffold. In Experiment 1, counterfactuals support evidence evaluation, leading children to privilege and extend the cause that accounted for more data. In Experiment 2, the hypothesis that accounted for the most evidence was pitted against children s prior beliefs. Children who considered alternative outcomes privileged the hypothesis that accounted for more observations, whereas those who explained relied on prior beliefs. Findings demonstrate that counterfactuals recruit attention to disambiguating evidence and outperform explanation when data contrast with existing beliefs.

Young learners have remarkable causal reasoning abilities. Children make accurate causal inferences from patterns of incoming data (Gopnik, Sobel, Schulz, & Glymour, 2001) and revise their hypotheses in light of new evidence (Schulz, Bonawitz, & Grif ths, 2007). However, it is often the case that more than one hypothesis is consistent with a learner s observations. Imagine, for instance, that a child observes muddy paw prints and puddles leading from the door, across the kitchen floor. She may infer that (a) it rained last night or (b) the sprinkler system came on.

When evidence is inconclusive, prior work shows that preschoolers tend to seek additional information (Gweon & Schulz, 2008; Schulz & Bonawitz, 2007) and take actions to discriminate between competing hypotheses (Cook, Goodman, & Schulz, 2011; Lapidow & Walker, 2020; Sobel & Sommerville, 2010). For example, preschoolers are more likely to explore a toy after observing a confounded event (e.g., an experimenter pulling two levers on the toy simultaneously, causing two objects to appear) than an unconfounded one (e.g., an experimenter pulling each lever on the toy in turn, causing one object to appear at a time; Schulz & Bonawitz, 2007). Children will also spontaneously intervene to isolate candidate causes to test their individual effects (e.g., separating beads that were snapped together) when the evidence they observe is ambiguous (i.e., when they know that only *some* beads are causal), and not otherwise (i.e., when they know that *all* beads are causal; Cook et al., 2011).

In contrast, other work suggests that young children may fail to discriminate between competing hypotheses when asked to make an explicit causal judgment. For example, Walker et al. (2016) presented 5-year olds with patterns of data (blocks that did or did not activate a machine), which suggested two possible causal features (differently colored parts) that were equally consistent with children s prior knowledge. One feature accounted for 100% of their observations (e.g., all of the blocks with a red part activated the machine), and the other only accounted for 75% (e.g., 75% of the blocks with a white part activated the machine). Five-year olds failed to discriminate between the two candidate causes at test. Instead, children treated both causes equally when they were pit against one another (i.e., red parts vs. white parts), failing to privilege

We are grateful to the children and parents who participated in the study, as well as the Reuben H. Fleet Science Museum for their help with recruitment. We also thank Hazel Baker-Harvey, Andie Nishimi, Nicky Sullivan, and Angela Wu for their efforts toward data collection. We thank Leslie Carver, Patricia Ganea, and Angela Nyhout for their feedback in conceptualizing this project. This work was partly supported by a Hellman Fellowship to Caren M. Walker.

Correspondence concerning this article should be addressed to Jae Engle, Department of Psychology, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093. Electronic mail may be sent to jengle@ucsd.edu.

 0009-3920/2021/xxxx-xxxx
DOI: 10.1111/cdev.13518

the 100% feature when making a causal judgment about a novel case.

In an effort to improve this ability, some researchers have relied on specific scaffolds. In the study described earlier, Walker et al. (2016) demonstrated that a prompt to explain facilitates this judgment: Children who were asked to generate explanations for the evidence they observed were more likely to privilege the cause that better accounted for the data. Indeed, similar results have been found in adults (Williams & Lombrozo, 2013; Williams, Lombrozo, & Rehder, 2013). Critically, however, prompts to explain are not always beneficial. Specifically, because explaining tends to recruit the learner's attention to *specific types* of hypotheses—that is, those that are broad, generalizable, simple, and abstract (Lombrozo, 2012; Walker, Bonawitz, & Lombrozo, 2017; Walker & Lombrozo, 2017)—the act of generating an explanation can sometimes lead learners astray.

For instance, Williams and Lombrozo (2013) discovered that prompts to explain sometimes cause adult learners to commit errors of *over*generalization, dismissing exceptions to a broad or abstract rule. In some cases, this leads learners to discount new evidence that is incompatible with their prior beliefs (e.g., Kuhn & Katz, 2009; Walker et al., 2016). Walker et al., (2016, exp. 3) found that when children's beliefs about likely causes (i.e., larger/heavier objects are more likely to be causal) conflicted with a hypothesis that better accounted for their current observations (i.e., block color), 5-year-old explainers were more likely than controls to select the more familiar hypothesis that better aligned with their existing knowledge. Therefore, while explanation can facilitate learning in instances where the correct hypothesis is broadly generalizable and in line with the learner's prior commitments, it can also mislead learners when a less familiar or an initially less likely hypothesis better accounts for the data. Here, we consider whether prompting learners to consider alternatives to a focal hypothesis by engaging in *counterfactual reasoning* may provide a more effective strategy than explanation to support causal learning under these circumstances.

Counterfactual Reasoning

Prior research provides evidence that counterfactual reasoning—comparing a real event to some alternative, imagined outcome—can impact learning and inference in both children and adults (e.g., Frosch, McCormack, Lagnado, & Burns, 2012; Guajardo & Turley-Ames, 2004; Harris, German, & Mills, 1996; McCormack, Simms, McGourty, & Beckers, 2013; Nyhout, Iannuzziello, Iannuzziello, Walker, & Ganea, 2019; Roese, 1994; Wells & Gavanski, 1989). This work falls under several broad theoretical accounts (see Nyhout & Ganea, 2019; Walker & Nyhout, 2020 for review).

First, it has been suggested that reasoning about a specific counterfactual may make the parallel causal inference more accessible to the learner (Byrne, 2005; Roese & Olson, 1997; Tetlock & Lebow, 2001). To illustrate, consider the following scenario: "Mary fails an exam and then muses that she would have passed had she worked through the study guide" (Roese & Olson, 1997, p. 142). Roese and Olson (1997) note that these counterfactuals are useful for learning since they help the reasoner to identify the critical causal antecedent (i.e., working through the study guide) that would influence the outcome of some event (e.g., passing or failing the exam), thus serving a preparative function by supporting future decision-making and causal attributions (Epstude & Roese, 2008; Roese, 1997).

Others have proposed that counterfactual reasoning leads to a *counterfactual mindset* (Epstude & Roese, 2008; Galinsky & Moskowitz, 2000; Kray & Galinsky, 2003), which has been shown to invoke an openness to alternative possibilities (i.e., a "simulation mindset"; Galinsky & Moskowitz, 2000), a tendency to consider the relationships between concepts and events (i.e., "relational processing"; Kray, Galinsky, & Wong, 2006), and a broadening of conceptual attention (i.e., "expansive processing"; Markman, Lindberg, Kray, & Galinsky, 2007). In one instance, adults who were primed with a narrative describing a "near-miss" counterfactual scenario (i.e., a character narrowly missing out on winning a prize) were more likely to seek evidence that would disconfirm an existing hypothesis on an unrelated task—that is, asking personality test questions to disconfirm their belief that a character is an extrovert (e.g., "What factors make it hard for you to open up to people?"), rather than to confirm this belief (e.g., "What do you like about parties?"; Galinsky & Moskowitz, 2000). These and related findings suggest that counterfactual reasoning may broadly serve to mitigate bias by directing the learner to consider alternatives to a focal hypothesis—in this case, that the character is actually an introvert. This ability to overcome the pull of an existing belief is of particular importance when a learner holds some prior theory that is incompatible with the new evidence they observe (e.g., Karmiloff-Smith & Inhelder, 1974; Schulz et al., 2007). It has been proposed that, in these cases, engaging in counterfactual reasoning could be particularly bene cial, leading the learner to elevate hypotheses that are initially considered lower probability or even counterintuitive (Walker & Nyhout, 2020). Notably, this also requires the learner to flexibly switch from an initial hypothesis, and there is prior evidence that individual differences in children s cognitive flexibility are related to their spontaneous generation of counterfactuals (Guajardo, McNally, & Wright, 2016).

A third broad account suggests that counterfactual reasoning is closely related to the formation of causal judgments and may impact the *process* of causal inference itself (Buchsbaum, Bridgers, Weisberg, & Gopnik, 2012; Gopnik & Walker, 2013; Harris et al., 1996; Mandel, 2003; Nyhout et al., 2019; Spellman, Kincannon, & Stose, 2005; Walker & Gopnik, 2013a; Woodward, 2007). Notably, the extent to which counterfactuals play a role in causal reasoning is the subject of some debate. Some argue that a distinguishing feature of causal knowledge is the fact that causal relations require counterfactual dependence (i.e., the causal relationship "*X causes Y*" implies the counterfactual, "*a change to X would lead to a change in Y*"; Gopnik, Schulz, & Schulz, 2007; Lewis, 1986; Mackie, 1974; Pearl, 2000; Woodward, 2003). From this perspective, counterfactuals serve as inputs to causal judgments by changing the value of a causal variable and considering the downstream effects of that change (i.e., an "imagined intervention"; Gopnik & Walker, 2013; Sloman, 2005; Walker & Gopnik, 2013a, 2013b).

Others have suggested that counterfactuals play a *speci c*, task-dependent role in causal inference, but stop short of interpreting these effects as evidence for a counterfactual theory of causation (e.g., Hoerl, McCormack, & Beck, 2011; McCormack et al., 2013). For example, McCormack et al. (2013) introduced 5- to 7-year olds to a novel causal system in which some "foods" caused a robot to play a sound and light up while others did not. When the experimenter provided *one* causal food, there was a small effect (faint sound and light), and when she provided *two* causal foods, there was an additive effect (ampli ed sound and light). Children were assigned to either report what happened to the robot (factual) or imagine what would happen if a non-causal food had been causal (counterfactual). At test, children who were asked to consider counterfactuals displayed enhanced "blocking" compared to children who were asked to consider factual questions, inferring that if two foods failed to produce an ampli ed effect, they could not both be causal. The authors proposed that counterfactual thinking serves to highlight the additive relationship between the causes and the effect produced, facilitating higher-order reasoning in young children.

Across accounts, counterfactual questions have been shown to facilitate learners ability to access and apply a variety of inferential reasoning skills. Here we extend this literature to explore whether counterfactual reasoning may also be applied to support children s evaluation of evidence. To examine this, we modi ed Walker et al. (2016) causal learning paradigm to assess whether counterfactual prompts will help children differentiate between candidate hypotheses that vary in their consistency with the evidence and prior beliefs. Although the prior theoretical accounts of counterfactual reasoning reviewed earlier provided motivation for the current study, we do not aim to distinguish among these proposals. Instead, the goal of the present work is to add to this growing body of developmental research to examine whether counterfactual reasoning meaningfully contributes to children s ability to discriminate between competing hypotheses during causal learning. However, we return to consider the contributions of the current work to existing theories of counterfactual reasoning in the discussion.

Following Walker et al. (2016), the current experiments focus on 5-year olds. We selected this age group for two reasons. First, 5-year olds possess the cognitive prerequisites needed for this experimental paradigm, which requires children to track information about more than one candidate cause. Speci cally, developmental studies suggest that by age 5, children can reason successfully in a variety of causal inference tasks (e.g., Gopnik & Sobel, 2000), are capable of engaging in probabilistic reasoning (e.g., Gopnik et al., 2004; Grif ths, Sobel, Tenenbaum, & Gopnik, 2011; Schulz et al., 2007), can form novel inferences based on co-variation data (Kushnir, Xu, & Wellman, 2010), and reflect on multiple possibilities (Beck, Robinson, Caroll, & Apperly, 2006). Additionally, as mentioned earlier, encouraging 5-year olds to think counterfactually leads them to engage in more sophisticated forms of causal inference (e.g., McCormack et al., 2013).

In the current studies, we test (a) whether considering alternatives helps learners discriminate between competing hypotheses that vary in their consistency with the evidence observed and (b) whether counterfactuals outperform prompts to explain when the evidence conflicts with prior commitments. We predict that, like explanation, counterfactual prompts will enhance children s ability to detect disambiguating data, leading them to privilege the cause that accounts for a greater proportion of their observations. We also anticipate that the benefits of counterfactual prompts will *continue to hold*, even in contexts in which prompts to explain lead to errors of overgeneralization in line with the learner's prior beliefs. Specifically, counterfactual prompts may mitigate learners' tendency to dismiss evidence that does not align with their focal hypothesis. In addition to examining the specific role of counterfactual prompts in supporting evidence evaluation in young children, these studies aim to shed new light on the specific cognitive mechanisms underlying counterfactual reasoning, differentiating its role from related scaffolds, and motivating future work.

Experiment 1

The causal learning procedure used in Experiment 1 was similar to Walker et al. (2016, exp. 1). Over the course of eight events, children observed a novel causal system in which some blocks made a toy light up (causal), and other blocks did not (inert). Each block was designed with two colored features: (a) a fully associated 100% feature and (b) a partially associated 75% feature. The 100% feature accounted for the causal efficacy of *all* blocks, whereas the 75% feature accounted for the causal efficacy of *most* blocks. Both were equally plausible with respect to children's prior beliefs, varying only in color (see Figure 1).

Half of the children in Experiment 1 were prompted to consider alternative outcomes (counterfactual condition), and the other half were prompted to describe the outcomes observed (control condition). In line with prior work, description was selected as a comparison task because it shares many commonalities with counterfactual prompts: it requires children to attend to the evidence they observe, to verbalize in a social context, and to spend an equivalent amount of time engaging with each causal event. In this experiment, we examine whether prompts to consider alternatives might facilitate children's ability to detect disambiguating data to support their capacity to discriminate between two competing hypotheses.

Method

Participants

We recruited a planned sample of sixty-four 5-year-olds (M = 65.0 months, SD = 3.97 months, range = 59.6–72.0; 36 females). Children were randomly assigned to *control* (n = 32) or *counterfactual* (n = 32) conditions, with no significant difference in age between groups, p = .311. An additional 12 children were tested, but excluded due to inattention (5), lack of comprehension (3), parental interference (1), or experimenter error (3). Children were recruited from local preschools, museums, and a university subject pool. While individual demographics were not collected, the population was predominately white (44.5%) and middle-class (median household income $73,900).

Materials

Machine. The machine used in the training phase was a "blicket detector" (Gopnik & Sobel, 2000), which consisted of a black (5 × 5 × 4¾) wooden box containing a light that was surreptitiously controlled by a remote. Certain blocks were said to cause the toy to light up when placed on top.

Training and observation blocks. A total of eight 3-in. wooden cubes were used (4 causal and 4 inert) in the *training* (2 blocks) and *observation* (6 blocks) *phases*. Only the causal blocks "activated" the toy, causing it to light up. Each block had a plastic Lego plate affixed to both the *top* and *front* of each cube. On each Lego plate, there was a small square Lego piece (red, white, blue, or yellow) attached. One of the two Lego colors represented the 100% feature, and the other Lego color represented the 75% feature. During the eight observations, one color would *always* correlate with the machine lighting up (100% of the time), and the other color would correlate *some* (75%) of the time (see Figure 1).

For example, consistent with the 100% feature, all four causal blocks might have a red Lego on top, and all four inert blocks might have a blue Lego on top. Consistent with the 75% feature, three out of the four *causal* blocks would also have a white Lego in front, and one would have a yellow Lego in front, and three out of the four *inert* blocks would also have a yellow Lego in front, and one would have a white Lego in front. The placement of the colors and the color representing the 100% causal feature were counterbalanced.

Two reminder cards were used to help children remember which blocks were causal and which were inert, minimizing cognitive load. The causal blocks were placed next to a card with a picture of the toy with the light on. The inert blocks were placed next to a second card with a picture of the toy with the light off.

Test blocks. Two novel blocks were used to test for generalization. One block included the 100% feature (e.g., a red Lego) plus a novel color (a purple Lego), and the other block included the 75% feature (e.g., a white Lego) plus another novel color (e.g., a green Lego).

Procedure

Training phase. The training phase served to familiarize children with the experiment by introducing the materials and the novel causal system. The experimenter brought out the machine and said,

> This is my toy. When I put some things on top of my toy, my toy will light up. When I put other things on top of my toy, it will not light up. Let s try to nd out what things will make my toy light up.

The experimenter then brought out the rst block and said, “Let s try this one,” and placed it on the toy. The rst block was always causal (e.g., red Lego on top, white Lego in front), the toy lit up, and the experimenter asked, “Did this one make my toy light up or not light up?” After providing a verbal response, children were asked to sort the block in front of the appropriate reminder card (causal or inert). This process was then repeated for an inert block (e.g., blue Lego on top, yellow Lego in front).

Observation phase. During the observation phase, participants were randomly assigned to *control* or *counterfactual* conditions. Children observed six trials in which the experimenter placed all of the remaining blocks on top of the toy.

After each demonstration, the child was asked two questions. The rst question was the same, regardless of condition: “Let s try this one ” Then, after the demonstration, the child was asked, “Did this one make my toy light up or not light up?” As in the training phase, the children were again asked to sort the block in front of either the causal or inert memory card. The second question then differed by condition. For instance, in the *control* condition, if the child observed a red/white block causing the toy to light up, they were asked,

> Now I want you to remind me what happened. What happened when I put this one (*points to the block)*, shown here (*shows a picture of the same block*) on top of my toy? Did my toy light up or not light up?

In the *counterfactual* condition, the experimenter asked,

> Now I want you to imagine something different. What if this one (*points to the block*) had been this (*shows a picture of a different block)?* What would have happened to my toy? Would my toy have lit up or not lit up?

The *counterfactual* prompt served to explicitly call attention to an alternative scenario: Whenever the experimenter referred to a causal block, they would ask the participant to imagine what would have happened if the block had included inert features. Similarly, whenever the experimenter referred to an inert block, they would ask the participant to imagine what would have happened if the block had included causal features (see Figure 1). In both cases, the participant was invited to consider an alternative *outcome*. All of the alternative blocks that participants were invited to imagine were from the training phase (red/white or blue/yellow); disambiguating blocks were not used in counterfactual prompts.

The presentation of the six blocks in the observation phase was pseudorandom (see Figure 1): Children rst observed two blocks that followed the same pattern of feature activation as the two used in the training phase (e.g., red/white [causal] and blue/yellow [inert]). Critically, the training phase presented confounded evidence consistent with *both* hypotheses (100% feature, 75% feature). Next, children observed two instances of disambiguating evidence: one block with a red Lego on top and a yellow Lego in front activating the toy (causal), and one block with a blue Lego on top and a white Lego in front not activating the toy (inert). These blocks provided the critical evidence needed to discriminate the two hypotheses: red, and not white, is the causal feature that accounts for more of the data. Finally, children observed two blocks that were again consistent with the original pattern of activation. All blocks remained sorted and in full view throughout the entire experiment both to reduce memory demands and to help children in both conditions integrate information received across trials.

Generalization phase. During the generalization phase, the experimenter said,

> Now that you ve seen how my toy works, I need your help nding more things that will make my toy light up. I have some more blocks inside of this bag. I m going to tell you about some of

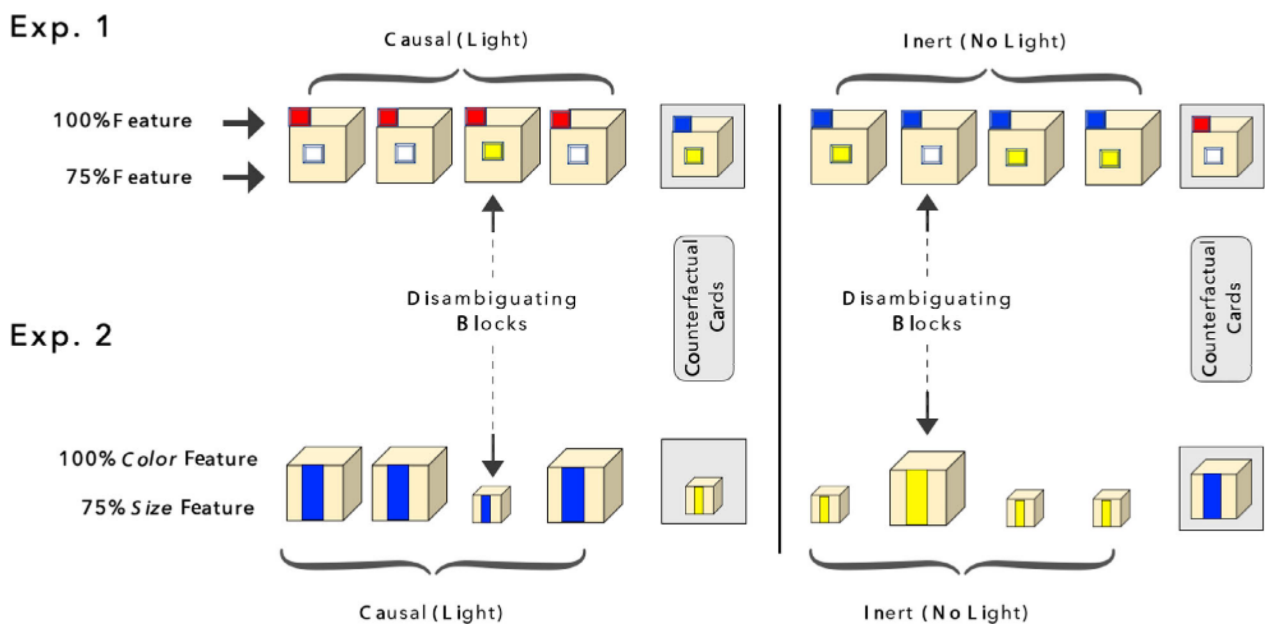

Figure 1. The set of causal and inert blocks used in Experiment 1 (top row) and Experiment 2 (bottom row), along with the counterfactual card depicting the alternative block for each set. All alternative blocks were from the training phase (blue/yellow or red/white for Experiment 1 and small/yellow or large/blue) for Experiment 2. In Experiment 1, the 100% feature is depicted by the red (causal) and blue (inert) Legos on top of the blocks. The 75% feature is depicted by the white (causal) and yellow (inert) Legos on the front of the blocks. In Experiment 2, the 100% *color* feature is depicted by the blue (causal) and yellow (inert) bands. The 75% *size* feature is depicted by relative size of blocks: large (causal) and small (inert). The placement of the causal features was counterbalanced across participants.

> these blocks and ask you some questions about them.

The experimenter then looked inside an opaque bag (the contents of which were not visible to the participant) and asked two *no-conflict* questions, followed by two *conflict* questions (*verbal conflict* and *transfer conflict*). In the *no-conflict* questions, the 100% *causal* feature was pitted against the 100% *inert* feature. Likewise, the 75% *causal* feature was pitted against the 75% *inert* feature. For example, in the *no-conflict* question, red was pitted against blue because red activated the toy 100% of the time, and blue was inert 100% of the time. The experimenter said,

> When I look inside this bag, I see one with a red part, and I see one with the blue part. Which one will make my toy light up the one with the blue part or the one with a red part?

The order of presentation was counterbalanced. These *no-conflict* questions were included to ensure that children in both conditions inferred the co-variation pattern between block color and activation of the machine.

For all *conflict* questions, the 100% feature (e.g., red) was pitted against the 75% feature (e.g., white) to assess which (if any) feature would be favored. The procedure for the *verbal conflict* question was identical to the *no-conflict* questions: The experimenter looked inside the bag and described features of the two blocks. However, in this case, the features were *both* associated with the effect, with varying probability (e.g., 100% red vs. 75% white).

Finally, in the *transfer conflict* question, the experimenter placed two novel blocks on the table in front of the child. The rst block was composed of the 100% feature (e.g., red), paired with a novel color (e.g., purple). The second block was composed of the 75% feature (e.g., white), paired with a different novel color (e.g., green). The experimenter said, "I have two new blocks. Which one will make my toy light up? The one with the red/purple part or the one with the green/white part?" This question not only assessed whether counterfactual prompts facilitated generalization to a novel case but it also served to address a potential ambiguity of the verbal test question. That is, if a child hears, "I have one with a red part and one with a white part," they may reasonably infer that the object with the white part also has a red part. Critically, to answer both types of *conflict* questions correctly, children must notice the disambiguating evidence and use this information to inform their subsequent inferences.

Results

Our planned analysis tested three hypotheses: (a) Children in both conditions will successfully learn the 100% and 75% features from co-variation data (*no conflict* questions). (b) Counterfactual prompts will facilitate children s ability to detect disambiguating evidence to discriminate between candidate causes (*conflict* questions). (c) Counterfactual prompts will help young learners generalize newly learned rules to a novel case (*transfer conflict* question).

No-Conflict Test Items

First, to test whether children learned the pattern of co-variation between the two colors and the toy lighting up, we conducted exact binomial tests. Average proportions of responses for all question types in each condition appear in Figure 2.

Children in both the *control* and *counterfactual* conditions performed signi cantly above chance on the 100% feature *no-conflict* question (*control*: $M = .96$, $SD = .18$; *counterfactual*: $M = 1$, $SD = 0$), $p < .001$, and 75% feature *no-conflict* question (*control*: $M = .81$, $SD = .40$; *counterfactual*: $M = .75$, $SD = .44$), $p < .01$, with no difference between conditions, $p = .76$. This indicates that children in both conditions successfully learned the novel causal structure and that counterfactual prompts did not simply increase general attention or engagement compared to controls.

Conflict Test Items

To test whether counterfactual prompts influenced children s ability to detect disambiguating data to make subsequent causal inferences, we next analyzed performance on both the *verbal* and *transfer conflict* questions. In line with our prediction, children in the *counterfactual* condition privileged the 100% feature ($M = .81$, $SD = .4$), $p < .001$ (exact binomial test), while those in the *control* condition selected between two options at chance ($M = .53$, $SD = .51$), $p = .13$. There was a signi cant difference between conditions, with children in the *counterfactual* condition more likely to privilege the 100% feature than those in the *control* condition, $^2(1) = 4.54$, $p = .03$ (two-tailed), $= .3$ (Fishers exact).

We found the same pattern of results for responses to the *transfer conflict* question, in which children in the *counterfactual* condition privileged the 100% feature ($M = .81$, $SD = .4$), $p < .001$ (exact binomial test), while those in the *control* condition selected between the two options at chance ($M = .53$, $SD = .51$), $p = .13$. Again, there was a signi cant difference between conditions, $^2(1) = 4.54$, $p = .03$ (two-tailed), $= .3$ (Fishers exact).

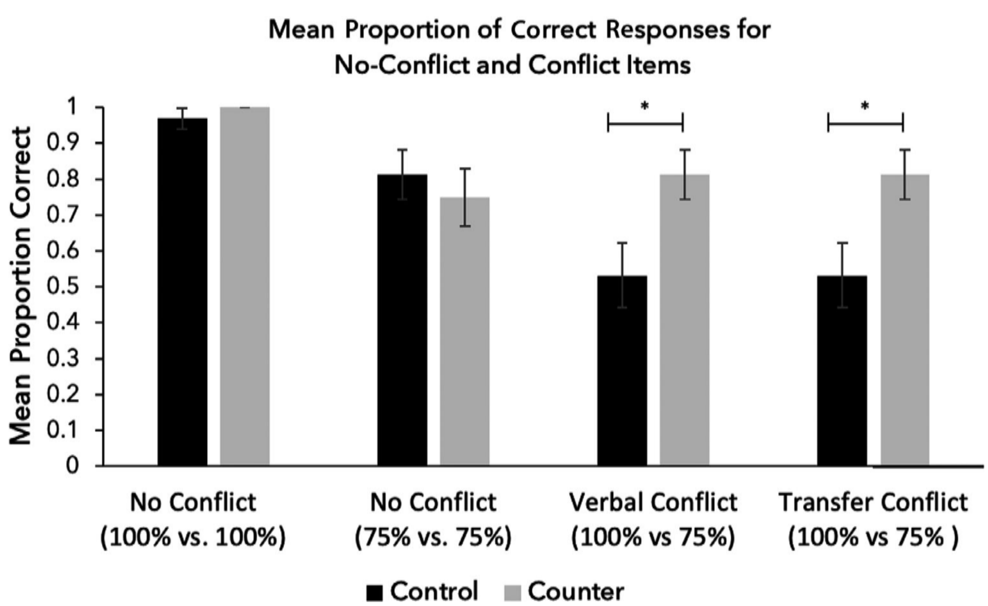

Figure 2. Mean proportion of correct responses for No-Conflict and Conflict items in *control* and *counterfactual* conditions.

Table 1
Accuracy of Counterfactual, Verbal, and Transfer Conflict Questions in Experiment 1

Number of counter-factual questions answered correctly	Number of participants	% accuracy on verbal question	% accuracy on transfer question
0			
1	2	100%	100%
2	3	66.6%	100%
3	2	100%	50%
4	1	100%	100%
5	5	100%	80%
6	19	73.7%	78.9%

Qualitative Analysis of Counterfactuals

Finally, we conducted an additional exploratory analysis to assess whether the accuracy of children s responses to the counterfactual questions predicted their performance on *verbal* and *transfer* conflict questions. To ensure that children had suf cient knowledge to answer the counterfactual questions, we only analyzed responses to counterfactual prompts for observations that occurred after the initial two training trials. Table 1 shows that while the majority of the participants answered all counterfactual questions correctly, those who did not provide consistent responses to these prompts still performed well on both the *verbal* and *transfer conflict* questions. This could suggest that the *process* of generating alternatives is suf cient to facilitate the detection of disambiguating data, regardless of the accuracy of children s responses. Indeed, a similar dissociation has been reported in the explanation literature (Walker, Lombrozo, Legare, & Gopnik, 2014; Walker et al., 2016). Although too few children provided incorrect responses to de nitively draw this conclusion for the current study, we will return to this point in our discussion of Experiment 2.

Discussion

Results from Experiment 1 suggest that prompting 5-year olds to consider alternatives facilitates detection of disambiguating evidence, increasing their tendency to privilege the candidate cause that is consistent with the greatest number of observations and transfer this newly learned rule to a novel case. Furthermore, the fact that children in both conditions successfully answered the *no-conflict* questions provides initial evidence against the alternative that counterfactual prompts simply boost overall attention or engagement during training trials. It is also unlikely that the counterfactual condition simply served to direct low-level perceptual attention to the critical evidence. First, although children must attend to the colors of the disambiguating block to imagine an alternative, this is true for all blocks, across trials. Second, despite drawing attention to the block s color, children must still infer *why* the color combination on the critical block disambiguates between the candidate causes.

Notably, these ndings are broadly consistent with previous reports on the effects of prompts to explain (Walker et al., 2016, exp. 1). However, as described earlier, we anticipate that counterfactual reasoning should outperform explanation when discriminating between competing hypotheses in cases where the evidence is inconsistent with children s prior beliefs about the most likely cause. We investigate this proposal in Experiment 2.

Experiment 2

Experiment 2 directly compares the effects of counterfactual and explanation prompts in a context in which explanation is known to lead learners astray. To do so, participants were again presented with a series of eight observations that were consistent with two candidate causes: a 100% feature that accounted for all of the evidence (color), and a 75% feature that was more consistent with children s prior beliefs (size). Size was selected based on prior work showing that children favor large/heavy objects as a more plausible causal factor than super cial properties, like color (Esterly, 2000; Walker et al., 2016). Critically, using the identical method, Walker et al. (2016, study 1) report that children frequently appealed to the weight of the blocks in their explanations, even though the blocks did not vary in weight (or size), suggesting that this hypothesis was driven by prior beliefs and not the evidence observed. The authors also veri ed this assumption with a computational model (study 3), demonstrating that children assign the highest prior probability to the hypothesis that larger blocks activate the machine.

Here, we predict that when these prior beliefs are available, counterfactual and explanation prompts will produce diverging results on causal judgments. Speci cally, we expect that children who are prompted to consider alternatives will not default to the hypothesis that is most in line with their prior knowledge at test. Instead, the process of considering counterfactuals should mitigate the likelihood that children will simply dismiss inconsistent evidence. Thus, we predict that children prompted with counterfactuals will be more likely than those prompted to explain to privilege the hypothesis that accounts for a greater proportion of the current data.

Method

Participants

We recruited a planned sample of fty-six 5-year olds (M = 65.45 months, SD = 4.23 months, range = 55.8 74.8; 26 females). Children were randomly assigned to *explanation* (n = 28) or *counterfactual* (n = 28) conditions, with no signi cant difference in age between groups, p = .60. An additional nine children were tested, but excluded due to inattention (1), parent interference (3), or experimenter error (5). Participant recruitment and demographics were identical to those in Experiment 1.

Materials

Experiment 2 used the same machine as Experiment 1. The causal blocks consisted of three 3-in. wooden cubes and one 1-in. wooden cube. The inert blocks consisted of three 1-in. wooden cubes and one 3-in. cube. For the 100% color feature, a half-inch piece of colored (e.g., blue or yellow) electrical tape was af xed around each of the eight blocks. The four causal blocks featured blue electrical tape, and the four inert blocks featured yellow electrical tape.

Consistent with the 75% size feature, three of the four causal blocks were large (3 in.) wooden cubes, and the remaining causal block was smaller (1 in.) wooden cube. The reverse pattern was used for the inert blocks (see Figure 1).

Procedure

Training/Observation phase

The training and observation phases were identical to those used in Experiment 1, except for the following changes:

First, participants were randomly assigned to *explanation* or *counterfactual* conditions. As in Experiment 1, all children observed eight trials in which the experimenter placed blocks on top of the toy (see Figure 1). Again, after each demonstration, the child was asked two questions. The rst question was the same, regardless of condition: "Let s try this one " Then, after the demonstration, they were asked, "Did this one make my toy light up or not light up?" and were instructed to sort the block in front of either the causal or inert reminder card.

The second question differed by condition. In the *explanation* condition, if the child observed a big (3 in.) wooden block with blue tape causing the toy to light up, the child was asked, "Now I want you to tell me something. Why did this one (*points to the block*), shown here (*shows a picture of the same block*) make my toy light up?" In the *counterfactual* condition, the experimenter asked,

> Now that you ve seen what s happened I want you to imagine something different. What if this one (*points to the block*) had been this (*shows a picture of a different block)?* Would my toy have lit up or not lit up?

The *counterfactual* prompt served to explicitly call attention to an alternative scenario. Similar to Experiment 1, the participant was always invited to imagine the opposite outcome: for any causal block, children were asked to consider what would happen if the block had included inert features, and for any inert block, they were asked what would happen if the block had included causal features (see Figure 1). Again, all of the alternative blocks were from the training phase (big/blue or small/yellow) and never presented a disambiguating block.

Test phase. The test phase was similar to the generalization phase in Experiment 1. However, given that we found no differences between the two *conflict* questions in Experiment 1 (p = 1), we only included a single conflict test question. The experimenter said, "I see two blocks. I see a big one, and I see a blue one. Which one will make my toy light up? The blue one or the big one?" This question similarly served to pit the two features against one another. To answer the *conflict* question correctly, children must notice the critical disambiguating evidence and override their prior beliefs that larger objects are more likely to be causal.

Results

Our planned analysis tested two hypotheses: (a) Children in both conditions will successfully learn the 100% and 75% features from co-variation data (*no conflict* questions). (b) Counterfactual prompts will help maintain children s sensitivity to the evidence in a context in which explanation prompts will lead them to rely more heavily on existing knowledge (*conflict* questions).

No-Conflict Test Items

To test whether children in both conditions learned the co-variation between block features and the toy lighting up, we conducted exact binomial tests. Average proportions of responses for all question types in each condition appear in Figure 3. As in Experiment 1, children in both conditions performed signi cantly above chance on the 100% color feature *no-conflict* questions (*explanation*: $M = .89$, $SD = .32$; *counterfactual*: $M = 1$, $SD = 0$), $p < .001$, and the 75% size feature *no-conflict* questions (*explanation*: $M = .89$, $SD = .32$; *counterfactual*: $M = .75$, $SD = .44$), $p < .01$, with no difference between conditions, $p = .30$. Again, children in both conditions successfully learned the novel causal structure, providing additional evidence that the conditions were well matched in terms of overall engagement and attention.

Conflict Test Item

To test whether prior beliefs influenced children s ability to detect disambiguating data to discriminate between competing hypotheses, we analyzed performance on the conflict question for both conditions. As in Experiment 1, and in line with our prediction, children in the *counterfactual* condition privileged the 100% color feature ($M = .89$, $SD = .31$), $p < .001$(exact binomial test). Critically, however, those in the *explanation* condition selected between two options at chance ($M = .57$, $SD = .50$), $p = .11$, replicating previous ndings (Walker et al., 2016, exp 3). This resulted in a signi cant difference between conditions, with children in the *counterfactual* condition more likely to privilege the 100% color feature over the 75% size feature that was more consistent with prior knowledge, $^2(1) = 5.83$, $p = .016$ (two-tailed), $= .36$ (Fishers exact). These ndings strongly support the claim that explanation and counterfactual prompts impact learning via distinct cognitive mechanisms.

Qualitative Analysis of Counterfactuals

Again, we conducted an exploratory analysis of the qualitative data. To ensure that children had suf cient knowledge to answer the counterfactual questions, we only analyzed responses to observations that occurred after the initial two training trials. Similar to Experiment 1, the majority of children (61%) answered all counterfactual questions accurately. Again, too few children responded incorrectly to determine whether the accuracy of children s responses to counterfactual questions predicted performance on the conflict question (see Table 2).

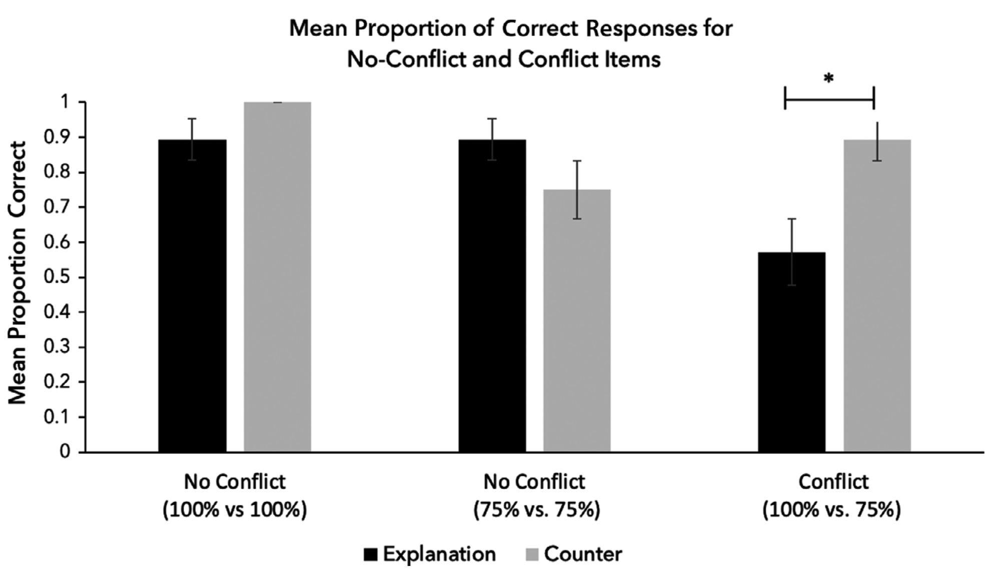

Figure 3. Mean proportion of correct responses for *No-Conflict* and *Conflict* items in *explanation* and *counterfactual* conditions.

Qualitative Analysis of Explanations

Most participants in the *explanation* condition (86%) provided explanations based on either color or size, with a minority of participants providing "other" explanations (e.g., "it's cold," "it doesn't work well," "the light is not coming"). Table 3 breaks down the distribution of participants' modal explanations as a function of their accuracy on the conflict question. Two children did not have a modal explanation and instead appealed equally to both size and color across the eight explanations provided.

Interestingly, many of the explanations generated were consistent with the evidence observed, emphasizing color over size. Although providing a modal explanation that was consistent with the 100% color feature led to a (non-significant) boost in accuracy, explainers were overall more likely to rely on prior knowledge than those prompted with counterfactuals when responding to the conflict question. One explanation for the discrepancy between the content of children's modal explanations and their performance at test is that the effect of generating explanations on subsequent causal judgments is dissociable from the specific explanations produced during learning. As noted earlier, similar findings have been reported in prior work, which points to the distinction between the *process* of explaining and the content of the explanations that are produced (Walker et al., 2014, 2016). For example, in one prior study, children who were asked to explain during learning exhibited more mature patterns of inference—privileging inductively-rich, but hidden features (e.g., causal relationships) over perceptually salient ones (e.g., color and shape)—even when the content of their explanations focused on superficial features (Walker et al., 2014).

Table 2
Accuracy of Counterfactual and Conflict Questions in Experiment 2

Number of counterfactual questions answered correctly	Number of participants	% accuracy on conflict question
0	1	100%
1		
2		
3	2	100%
4	3	100%
5	5	80%
6	17	88.2%

Table 3
Proportion of Correct Answers as a Function of Modal Explanation Type in Experiment 2

Explanation type	Number of participants	% accuracy on conflict question
Size/weight	7	42.9%
Color	17	64.7%
Other	2	50%
No modal explanation	2	50%

To account for these findings, Wilkenfeld and Lombrozo (2015) argue that the *act* of generating an explanation carries epistemic value, even when the learner fails to produce the correct explanation (i.e., "explaining for the best inference"). Specifically, they argue that searching for a good explanation triggers several beneficial processes, such as recruiting attention to causally relevant features (e.g., Legare & Lombrozo, 2014; Walker et al., 2014, 2016), encouraging comparisons between relevant cases (Williams & Lombrozo, 2013), and promoting abstraction (Walker et al., 2016; Williams & Lombrozo, 2010), each of which ultimately impacts children's subsequent causal judgments. Other accounts have suggested that when learners produce an incorrect explanation, it highlights the disconnect between the content of that explanation and the evidence (Chi, De Leeuw, Chiu, & LaVancher, 1994) or increases their metacognitive awareness of this gap (McNamara, 2004; Rozenblit & Keil, 2002), prompting greater understanding.

General Discussion

In two experiments, prompting children to consider alternative outcomes helped them to discriminate between competing hypotheses by improving their ability to detect disambiguating evidence. In Experiment 1, children observed patterns of evidence depicting two candidate causal features: (a) a fully associated 100% feature and (b) a partially associated 75% feature. When prior knowledge was held constant, the amount of evidence available for each cause was the only discernible difference between the two. In this context, children in the *counterfactual* condition privileged the 100% over the 75% feature and used it to generalize to a novel case, whereas children in the *control* condition failed to differentiate the two causes. These results mirror the effects of explanation in a similar paradigm

(Walker et al., 2016), suggesting that counterfactual prompts may provide a viable alternative to explanation as a means to encourage evidence evaluation in young children.

In Experiment 2, children observed a pattern of evidence in which the cause that accounted for all of the data was pitted against one that accounted for less, but was more consistent with their prior beliefs. In line with prior work (Walker et al., 2016), children who were prompted to explain were *less* likely to privilege the hypothesis that aligned with their current observations, indicating that they were relying more heavily on existing knowledge. In contrast, children who were asked to consider alternative outcomes remained sensitive to the presence of counterexamples, allowing them to privilege the hypothesis that accounted for more of their current observations. We can conclude, therefore, that although counterfactual reasoning and explanation are both "constructive" modes of learning (see Chi & Wylie, 2014) and are likely to be similarly engaging to children, they influence causal inference via distinct mechanisms. Speci cally, condition differences found in children s causal judgments suggest that considering alternatives may be a more appropriate scaffold than a prompt to explain when the evidence contrasts with an existing belief. That is, counterfactual prompts produce similar bene ts as explanation while avoiding some of the pitfalls.

The results of Experiment 2 also provide strong evidence against alternative, low-level explanations for the effects of counterfactual prompts on causal judgments. Speci cally, given that explanation prompts are similarly challenging, the bene ts of counterfactual thinking is unlikely to result from global effects on causal reasoning due to increased motivation, general attention, or depth of processing. However, it could still be argued that counterfactual prompts increased *directed* attention to disambiguating evidence by requiring that learners recall the outcomes of previous demonstrations. Could the process of retrieving this information support children s ability to integrate the evidence across trials? Features of the task design render this alternative explanation unlikely. Speci cally, to reduce memory demands, all blocks were grouped into categories with a reminder card indicating whether or not each one caused the machine to activate. The evidence for all trials remained sorted and in full view throughout the experiment, removing children s need to integrate information across demonstrations in both conditions.

Although the current studies were not designed to discriminate among existing accounts of counterfactual reasoning, our ndings build on and extend this prior work by establishing that prompts to consider alternative possibilities can help young learners evaluate evidence when making causal judgments. Speci cally, in addition to increasing attention to disambiguating evidence, counterfactual reasoning does not dampen children s sensitivity to hypotheses that may be initially less likely or unfamiliar. This pattern of results is broadly compatible with the predictions generated by all of the theoretical perspectives reviewed earlier, although in each case, there are also important differences between the prior work and the current methodology.

First, in line with claims that counterfactuals impact learning by making parallel causal inferences more accessible (e.g., Byrne, 2005; Roese & Olson, 1997; Tetlock & Lebow, 2001), the prompts used in the current study may have recruited children s attention to disambiguating evidence by encouraging them to consider the effects of *both* causal and inert features. According to Byrne (2011), reasoning counterfactually requires the learner to simultaneously hold two possibilities in mind the real event and the alternative which provides explicit access to previously implicit causal knowledge. Prior work with adults is consistent with this claim, demonstrating that counterfactual thinking guides future behavior by helping the learner to identify critical causal antecedents (Roese & Olson, 1997). Our ndings similarly suggest that counterfactuals support causal judgments in children, which not only inform future inferences but also facilitate generalization to novel cases.

Our results also align well with prior accounts suggesting that counterfactuals invoke an openness to alternative possibilities (Galinsky & Moskowitz, 2000; Hirt, Kardes, & Markman, 2004; Kray & Galinsky, 2003; Markman et al., 2007). These accounts would predict that reasoning counterfactually may have encouraged children to think more broadly and flexibly about the evidence, overriding the tendency to default to the most salient hypothesis. However, unlike this prior work, which used unrelated counterfactuals to prime a broad "counterfactual mindset," the prompts used in the current study were relatively narrow in scope, emphasizing speci c alternative outcomes for observed events. Therefore, although these prompts may have facilitated children s ability to engage with an alternative hypothesis, we cannot interpret our ndings as resulting from a shift in cognitive style or mode of processing. Future work should explore whether counterfactuals might produce more global effects on the process of evidence evaluation in children s causal learning.

Lastly, our results are also broadly consistent with predictions made by counterfactual theories of causal inference (e.g., Buchsbaum et al., 2012; Gopnik, 2009; Gopnik & Walker, 2013; Sloman, 2005; Walker & Gopnik, 2013a, 2013b; Woodward, 2007). As noted earlier, this account predicts that counterfactuals prompt children to consider the outcome of imagined changes to variables in the causal system. The current results add to this literature by suggesting that counterfactual thinking might facilitate causal inference by helping children to discriminate among competing candidate causes. When considered alongside prior work demonstrating that counterfactual thinking improves "blocking" in children s causal inferences (McCormack et al., 2013), these results may be interpreted as providing increasing support for this model. However, future work is needed to explain the discrepancies between these ndings and prior research that suggests a rather tenuous relationship between counterfactual and causal reasoning (e.g., Frosch et al., 2012; German, 1999; McCormack, Frosch, & Burns, 2011). One possibility is that a learner s ability to make appropriate counterfactual inferences relies, at least in part, on the speci c causal knowledge they possess in that domain (Sobel, 2011).

Although the current ndings do not provide suf cient evidence to distinguish among these prior accounts, they do provide a fruitful starting point to motivate future work on the effects of counterfactual reasoning on evidence evaluation. First, although the current study presented children with a relatively "lean" learning context, future work might consider how various contextual factors may mediate the effects of counterfactual reasoning on discriminating candidate causes. For example, prior work shows that counterfactual reasoning is closely related to a variety of social-cognitive variables, including curiosity (FitzGibbon, Moll, Carboni, Lee, & Dehghani, 2019), the valence of an outcome (e.g., German, 1999; Mandel, 2003), the learner s emotional state (e.g., regret and relief; McCormack, O Connor, Cherry, Beck, & Feeney, 2019; McMullen & Markman, 2000; Weisberg & Beck, 2012), and their theory of mind (e.g., Guajardo & Turley-Ames, 2004; Riggs & Peterson, 2000; Riggs, Peterson, Robinson, & Mitchell, 1998). Thus, there remain a variety of open questions regarding how these variables may interact with the current ndings. Might invoking children s curiosity about alternative outcomes have similar bene ts as explicit counterfactual questions? Given that children are more likely to engage in counterfactual thinking following a negative event (German, 1999), might manipulating the valence of the outcome change causal judgments?

To conclude, the current research extends existing evidence that counterfactual thinking facilitates causal reasoning, and demonstrates, for the rst time, the influence of these prompts on children s evidence evaluation. These ndings also reveal distinct cognitive mechanisms underlying counterfactual and explanation scaffolds, supporting prior claims that different types of prompts have unique and selective bene ts for learning.

References

Beck, S. R., Robinson, E. J., Caroll, D. J., & Apperly, I. A. (2006). Children s thinking about counterfactuals and future hypotheticals as possibilities. *Child Development*, *77*, 413 426. https://doi.org/10.1111/j.1467-8624.2006.00879.x

Buchsbaum, D., Bridgers, S., Weisberg, D. S., & Gopnik, A. (2012). The power of possibility: Causal learning, counterfactual reasoning, and pretend play. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*, 2202 2212. https://doi.org/10.1098/rstb.2012.0122

Byrne, R. M. (2005). *The rational imagination: How people create alternatives to reality*. Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/5756.003.0002

Byrne, R. M. J. (2011). Counterfactual and causal thoughts about exceptional events. In C. Hoerl, T. McCormack, & S. Beck (Eds.), *Understanding counterfactuals, understanding causation: Issues in philosophy and psychology* (pp. 208 229). Oxford UK: Oxford University Press.

Chi, M. T., De Leeuw, N., Chiu, M. H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, *18*, 439 477. https://doi.org/10.1016/0364-0213(94)90016-7

Chi, M. T., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, *49*, 219 243. https://doi.org/10.1080/00461520.2014.965823

Cook, C., Goodman, N. D., & Schulz, L. E. (2011). Where science starts: Spontaneous experiments in preschoolers exploratory play. *Cognition*, *120*, 341 349. https://doi.org/10.1016/j.cognition.2011.03.003

Epstude, K., & Roese, N. J. (2008). The functional theory of counterfactual thinking. *Personality and Social Psychology Review*, *12*, 168 192. https://doi.org/10.1177/1088868308316091

Esterly, J. B. (2000). *It s Size is Really Wood : The development of buoyancy and material kind understanding in children between three and seven years of age*. Berkeley, CA: University of California.

FitzGibbon, L., Moll, H., Carboni, J., Lee, R., & Dehghani, M. (2019). Counterfactual curiosity in preschool children. *Journal of Experimental Child Psychology*, *183*, 146 157. https://doi.org/10.1016/j.jecp.2018.11.022

Frosch, C. A., McCormack, T., Lagnado, D. A., & Burns, P. (2012). Are causal structure and intervention judgments inextricably linked? A developmental study. *Cognitive Science*, *36*, 261 285. https://doi.org/10.1111/j.1551-6709.2011.01208.x

Galinsky, A. D., & Moskowitz, G. B. (2000). Counterfactuals as behavioral primes: Priming the simulation heuristic and consideration of alternatives. *Journal of Experimental Social Psychology*, *36*, 384 409. https://doi.org/10.1006/jesp.1999.1409

German, T. P. (1999). Children s causal reasoning: Counterfactual thinking occurs for 'negative outcomes only. *Developmental Science*, *2*, 442 457. https://doi.org/10.1111/1467-7687.00088

Gopnik, A. (2009). *The philosophical baby: What children s minds tell us about truth, love & the meaning of life*. New York, NY: Random House.

Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L. E., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: causal maps and Bayes nets. *Psychological Review*, *111*(1), 3. https://doi.org/10.1037/0033-295x.111.1.3

Gopnik, A., Schulz, L., & Schulz, L. E. (Eds.). (2007). *Causal learning: Psychology, philosophy, and computation*. Oxford, UK: Oxford University Press.

Gopnik, A., & Sobel D. M. (2000). Detecting blickets: How young children use information about novel causal powers in categorization and induction. *Child Development*, *71*, 1205 1222. https://doi.org/10.1111/1467-8624.00224

Gopnik, A., Sobel, D. M., Schulz, L. E., & Glymour, C. (2001). Causal learning mechanisms in very young children: Two-, three-, and four-year-olds infer causal relations from patterns of variation and covariation. *Developmental Psychology*, *37*, 620. https://doi.org/10.1037/0012-1649.37.5.620

Gopnik, A., & Walker, C. M. (2013). Considering counterfactuals: The relationship between causal learning and pretend play. *American Journal of Play*, *6*(1), 15 28.

Grif ths, T. L., Sobel, D. M., Tenenbaum, J. B., & Gopnik, A. (2011). Bayes and blickets: Effects of knowledge on causal induction in children and adults. *Cognitive Science*, *35*, 1407 1455. https://doi.org/10.1111/j.1551-6709.2011.01203.x

Guajardo, N. R., McNally, L. F., & Wright, A. (2016). Children s spontaneous counterfactuals: The roles of valence, expectancy, and cognitive flexibility. *Journal of Experimental Child Psychology*, *146*, 79 94. https://doi.org/10.1016/j.jecp.2016.01.009

Guajardo, N. R., & Turley-Ames, K. J. (2004). Preschoolers generation of different types of counterfactual statements and theory of mind understanding. *Cognitive Development*, *19*(1), 53 80. https://doi.org/10.1016/j.cogdev.2003.09.002

Gweon, H., & Schulz, L. (2008, July). *Stretching to learn: Ambiguous evidence and variability in preschoolers exploratory play*. In Proceedings of the 30th annual meeting of the Cognitive Science Society (pp. 570 574)

Harris, P. L., German, T., & Mills, P. (1996). Children s use of counterfactual thinking in causal reasoning. *Cognition*, *61*, 233 259. https://doi.org/10.1016/s0010-0277(96)00715-9

Hirt, E. R., Kardes, F. R., & Markman, K. D. (2004). Activating a mental simulation mind-set through generation of alternatives: Implications for debiasing in related and unrelated domains. *Journal of Experimental Social Psychology*, *40*, 374 383. https://doi.org/10.1016/j.jesp.2003.07.009

Hoerl, C., McCormack, T., & Beck, S. R. (Eds.). (2011). *Understanding counterfactuals, understanding causation: Issues in philosophy and psychology*. Oxford, UK: Oxford University Press.

Karmiloff-Smith, A., & Inhelder, B. (1974). If you want to get ahead, get a theory. *Cognition*, *3*, 195 212. https://doi.org/10.1016/0010-0277(74)90008-0

Kray, L. J., & Galinsky, A. D. (2003). The debiasing effect of counterfactual mind-sets: Increasing the search for discon rmatory information in group decisions. *Organizational Behavior and Human Decision Processes*, *91*(1), 69 81. https://doi.org/10.1016/s0749-5978(02)00534-4

Kray, L. J., Galinsky, A. D., & Wong, E. M. (2006). Thinking within the box: The relational processing style elicited by counterfactual mind-sets. *Journal of Personality and Social Psychology*, *91*(1), 33. https://doi.org/10.1037/0022-3514.91.1.33

Kuhn, D., & Katz, J. (2009). Are self-explanations always bene cial? *Journal of Experimental Child Psychology*, *103*, 386 394. https://doi.org/10.1016/j.jecp.2009.03.003

Kushnir, T., Xu, F., & Wellman, H. M. (2010). Young children use statistical sampling to infer the preferences of other people. *Psychological Science*, *21*, 1134 1140. https://doi.org/10.1177/0956797610376652

Lapidow, E., & Walker, C. M. (2020). Informative experimentation in intuitive science: Children select and learn from their own causal interventions. *Cognition*, *201*, 104315. https://doi.org/10.1016/j.cognition.2020.104315

Legare, C., & Lombrozo, T. (2014). The selective bene ts of explanation on learning in early childhood. *Journal of Experimental Child Psychology*, *126*, 198 212. https://doi.org/10.1016/j.jecp.2014.03.001

Lewis, D. (1986). *Counterfactuals*. Cambridge, MA: Harvard University Press.

Lombrozo, T. (2012). Explanation and abductive inference. In *Oxford handbook of thinking and reasoning* (pp. 260 276). https://doi.org/10.1093/oxfordhb/9780199734689.013.0014

Mackie, J. L. (1974). Truth, probability, and paradox: A reply to James E. Tomberlin s review. *Philosophy and Phenomenological Research*, *34*, 593 594. https://doi.org/10.2307/2106821

Mandel, D. (2003). Counterfactuals, emotions, and context. *Cognition and Emotion*, *17*, 139 159. https://doi.org/10.1080/02699930302275

Markman, K. D., Lindberg, M. J., Kray, L. J., & Galinsky, A. D. (2007). Implications of counterfactual structure for creative generation and analytical problem solving. *Personality and Social Psychology Bulletin*, *33*, 312 324. https://doi.org/10.1177/0146167206296106

McCormack, T., Frosch, C., & Burns, P. (2011). The relationship between children s causal and counterfactual judgments. In C. Hoerl, T. McCormack, & S. R. Beck (Eds.), *Understanding counterfactuals, understanding causation: Issues in philosophy and psychology* (pp. 54 74). Oxford, UK: Oxford University Press.

McCormack, T., O Connor, E., Cherry, J., Beck, S. R., & Feeney, A. (2019). Experiencing regret about a choice helps children learn to delay grati cation. *Journal of Experimental Child Psychology*, *179*, 162 175. https://doi.org/10.1016/j.jecp.2018.11.005

McCormack, T., Simms, V., McGourty, J., & Beckers, T. (2013). Encouraging children to think counterfactually enhances blocking in a causal learning task. *Quarterly Journal of Experimental Psychology*, *66*, 1910 1926. https://doi.org/10.1080/17470218.2013.767847

McMullen, M. N., & Markman, K. D. (2000). Downward counterfactuals and motivation: The wake-up call and the Pangloss effect. *Personality and Social Psychology Bulletin*, *26*, 575 584. https://doi.org/10.1177/0146167200267005

McNamara, D. S. (2004). SERT: Self-explanation reading training. *Discourse Processes*, *38*(1), 1 30. https://doi.org/10.1207/s15326950dp3801_1

Nyhout, A., & Ganea, P. (2019). The development of the counterfactual imagination. *Child Development Perspectives*, *13*, 254 259. https://doi.org/10.1111/cdep.12348

Nyhout, A., Iannuzziello, A., Walker, C. M., & Ganea, P. (2019). Thinking counterfactually supports children s ability to conduct a controlled test of a hypothesis. In A. Goel, C. Seifert, & C. Freska (Eds.), *Proceedings of the 41st annual meeting of the Cognitive Science Society*, Montreal, CA.

Pearl, J. (2000). Causal inference without counterfactuals: Comment. *Journal of the American Statistical Association*, *95*, 428 431. https://doi.org/10.2307/2669380

Riggs, K. J., & Peterson, D. M. (2000). Counterfactual and mental state inferences in pre-school children. In I. N. P. Mitchell & K. J. Riggs (Eds.), *Children s reasoning and the mind* (pp. 87 99). Hove, UK: Psychology Press.

Riggs, K. J., Peterson, D. M., Robinson, E. J., & Mitchell, P. (1998). Are errors in false belief tasks symptomatic of a broader dif culty with counterfactuality? *Cognitive Development*, *13*(1), 73 90. https://doi.org/10.1016/s0885-2014(98)90021-1

Roese, N. J. (1994). The functional basis of counterfactual thinking. *Journal of Personality and Social Psychology*, *66*, 805. https://doi.org/10.1037/0022-3514.66.5.805

Roese, N. J. (1997). Counterfactual thinking. *Psychological Bulletin*, *121*, 133 . https://doi.org/10.1037/0033-2909.121.1.133

Roese, N. J., & Olson, J. M. (1997). Counterfactual thinking: The intersection of affect and function. In J. Greenberg, S. Solomon, & T. Pyszczynski (Eds.), *Advances in experimental social psychology* (Vol. *29*, pp. 1 59). San Diego, CA: Academic Press.

Rozenblit, L., & Keil, F. (2002). The misunderstood limits of folk science: An illusion of explanatory depth. *Cognitive Science*, *26*, 521 562. https://doi.org/10.1207/s15516709cog2605_1

Schulz, L. E., & Bonawitz, E. B. (2007). Serious fun: Preschoolers engage in more exploratory play when evidence is confounded. *Developmental Psychology*, *43*, 1045. https://doi.org/10.1037/0012-1649.43.4.1045

Schulz, L. E., Bonawitz, E. B., & Grif ths, T. L. (2007). Can being scared cause tummy aches? Naive theories, ambiguous evidence, and preschoolers causal inferences. *Developmental Psychology*, *43*, 1124. https://doi.org/10.1037/0012-1649.43.5.1124

Sloman, S. (2005). *Causal models: How people think about the world and its alternatives*. New York, NY: Oxford University Press.

Sobel, D. M. (2011). Domain-speci c causal knowledge and children s reasoning about possibility. In C. Hoerl, T. McCormack, & S. R. Beck (Eds.), *Understanding counterfactuals, understanding causation: Issues in philosophy and psychology* (pp. 123 146). Oxford, UK: Oxford University Press.

Sobel, D., & Sommerville, J. (2010). The importance of discovery in children s causal learning from interventions. *Frontiers in Psychology*, *1*, 176. https://doi.org/10.3389/fpsyg.2010.00176

Spellman, B. A., Kincannon, A., & Stose, S. (2005). The relation between counterfactual and causal reasoning. In D. R. Mandel, D. J. Hilton, & P. Catellani (Eds.), *The psychology of counterfactual thinking* (pp. 28 43). London, UK: Routledge.

Tetlock, P. E., & Lebow, R. N. (2001). Poking counterfactual holes in covering laws: Cognitive styles and historical reasoning. *American Political Science Review*, *95*, 829 843. https://doi.org/10.1017/s0003055400400043

Walker, C. M., Bonawitz, E., & Lombrozo, T. (2017). Effects of explaining on children s preference for simpler hypotheses. *Psychonomic Bulletin & Review*, *24*, 1538 1547. https://doi.org/10.3758/s13423-016-1144-0

Walker, C. M., & Gopnik, A. (2013a). Causality and imagination. In M. Taylor (Ed.), *The oxford handbook of the development of the imagination* (pp. 342 358). New York, NY: Oxford University Press.

Walker, C. M., & Gopnik, A. (2013b). Pretense and possibility A theoretical proposal about the effects of pretend play on development: Comment on Lillard et al (2013). *Psychological Bulletin*, *139*(1), 40 44. https://doi.org/10.1037/a0030151

Walker, C. M., & Lombrozo, T. (2017). Explaining the moral of the story. *Cognition*, *167*, 266 281. https://doi.org/10.1016/j.cognition.2016.11.007

Walker, C. M., Lombrozo, T., Legare, C. H., & Gopnik, A. (2014). Explaining prompts children to privilege inductively rich properties. *Cognition*, *133*, 343 357. https://doi.org/10.1016/j.cognition.2014.07.008

Walker, C. M., Lombrozo, T., Williams, J. J., Rafferty, A. N., & Gopnik, A. (2016). Explaining constrains causal learning in childhood. *Child Development*, *88*, 229 246. https://doi.org/10.1111/cdev.12590
Walker, C. M., & Nyhout, A. (2020). *13 Asking "Why?" and "What If?". The questioning child: Insights from psychology and education*. https://doi.org/10.1017/9781108553803.013
Weisberg, D. P., & Beck, S. R. (2012). The development of children s regret and relief. *Cognition & Emotion*, *26*, 820 835. https://doi.org/10.1080/02699931.2011.621933
Wells, G. L., & Gavanski, I. (1989). Mental simulation of causality. *Journal of Personality and Social Psychology*, *56*, 161. https://doi.org/10.1037/0022-3514.56.2.161
Wilkenfeld, D. A., & Lombrozo, T. (2015). Inference to the best explanation (IBE) versus explaining for the best inference (EBI). *Science & Education*, *24*, 1059 1077. https://doi.org/10.1007/s11191-015-9784-4
Williams, J. J., & Lombrozo T. (2010). The role of explanation in discovery and generalization: Evidence from category learning. *Cognitive Science*, *34*, 776 806. https://doi.org/10.1111/j.1551-6709.2010.01113.x
Williams, J. J., & Lombrozo, T. (2013). Explanation and prior knowledge interact to guide learning. *Cognitive Psychology*, *66*(1), 55 84. https://doi.org/10.1016/j.cogpsych.2012.09.002
Williams, J. J., Lombrozo, T., & Rehder, B. (2013). The hazards of explanation: Overgeneralization in the face of exceptions. *Journal of Experimental Psychology: General*, *142*, 1006. https://doi.org/10.1037/a0030996
Woodward, J. (2003). *Making things happen: A theory of causal explanation*. Oxford, UK: Oxford University Press.
Woodward, J. (2007). Interventionist theories of causation in psychological perspective. In *Causal learning: Psychology, philosophy, and computation* (pp. 19 36). https://doi.org/10.1093/acprof:oso/9780195176803.003.0002