CHILD DEVELOPMENT

Child Development, xxxx 2017, Volume 00, Number 0, Pages 1–13

Children's Counterfactual Reasoning About Causally Overdetermined Events

Angela Nyhout
University of Toronto

Lena Henke
University of Osnabrück

Patricia A. Ganea
University of Toronto

In two experiments, one hundred and sixty-two 6- to 8-year-olds were asked to reason counterfactually about events with different causal structures. All events involved overdetermined outcomes in which two different causal events led to the same outcome. In Experiment 1, children heard stories with either an ambiguous causal relation between events or causally unrelated events. Children in the causally unrelated version performed better than chance and better than those in the ambiguous condition. In Experiment 2, children heard stories in which antecedent events were causally connected or causally disconnected. Eight-year-olds performed above chance in both conditions, whereas 6-year-olds performed above chance only in the connected condition. This work provides the first evidence that children can reason counterfactually in causally overdetermined contexts by age 8.

The ability to think counterfactually is an important cognitive achievement. When thinking counterfactually, one considers how events, from the mundane to the life changing, could have turned out differently (e.g., "I wouldn't have spilled my juice if I hadn't been running."; "My children never would have been born if I hadn't met my partner."). It allows one to explain events in the past and make plans to adapt one's behavior in the future (Byrne, 2016; Epstude & Roese, 2008). To generate counterfactual inferences, reasoners take their representation of reality and make manipulations to it, setting forth a counterfactual premise and following its causal implications. In the present experiments, we investigated the different causal event structures children can manipulate counterfactually. We highlight the importance of correctly characterizing the starting point of this process—the representations of reality off which counterfactual inferences are based—in order to arrive at a response that is deemed correct by adults. We argue that children's apparent failures of counterfactual reasoning in previous work may instead have been due to mischaracterizations of their representations of reality.

As may be clear to the reader, the abilities to think causally and think counterfactually are closely intertwined. Reasoning about counterfactual scenarios often depends on importing relevant background knowledge about causes in order to draw correct counterfactual inferences (Sobel, 2011). Other researchers argue that drawing causal inferences may depend on reasoning about counterfactuals. On one view, arriving at a causal inference implies first reasoning counterfactually about what might have occurred but did not by comparing an observed sequence of events to an imagined sequence in which an event happened differently (e.g., Harris, German, & Mills, 1996; Lewis, 1973; Mackie, 1974). Other theories of the relation between causal and counterfactual reasoning assert that, when an individual makes a causal inference (e.g., X causes Y), they commit to the idea of counterfactual dependency (e.g., a change to X would lead to a change to Y; e.g., Gopnik & Schulz, 2007; Pearl, 2000; Schulz, Gopnik, & Glymour, 2007; Woodward, 2003). On this view, an individual who

This research was funded by an Insight Grant from the Social Sciences and Humanities Research Council of Canada and an Early Researcher Award from the Ontario Ministry of Research and Innovation to Patricia A. Ganea. We are grateful to the children and parents in the Greater Toronto Area who participated in this research. We thank Krisma Saini for assisting with data collection, and Maggie Hong, Lynn Nguyen, and Hilary Sweatman for assistance with coding. We also thank Brian Leahy for helpful feedback on an earlier version of this article.

Correspondence concerning this article should be addressed to Angela Nyhout, Dr. Eric Jackman Institute of Child Study, University of Toronto, 252 Bloor St West, Toronto, ON M5S 1V6. Electronic mail may be sent to angela.nyhout@utoronto.ca.

 0009-3920/2017/xxxx-xxxx
DOI: 10.1111/cdev.12913

has inferred a particular causal model should be able to make correct counterfactual predictions about interventions to variables in the model.

From a young age, children show an impressive ability to reason causally about events that did not happen but could have happened. For example, after being presented with a story about a character who left her muddy shoes on and made the kitchen floor all dirty, 3.5-year-olds were able to correctly answer that if she had removed her shoes, the floor would instead be clean (Harris et al., 1996). In another study, 3- and 4-year-olds were presented with a story in which a character planted a flower in her garden and called her husband to come see it, which allowed their dog to escape from the house and trample on the flower, making the character sad (German & Nichols, 2003). By age 4, children were able to correctly determine that the character would still be happy if she had not called her husband, indicating that they represented the causal chain of events when reasoning from the counterfactual. However, Beck, Riggs, and Gorniak (2010) attempted to replicate this result but did not find evidence that 4-year-olds were able to reason about causal chains in counterfactuals. Several other studies have found competence on counterfactual reasoning tasks by children's fourth birthdays (e.g., Beck, Robinson, Carroll, & Apperly, 2006, standard counterfactuals; Guajardo & Turley-Ames, 2004; Perner, Sprung, & Steinkogler, 2004; Riggs, Peterson, Robinson, & Mitchell, 1998; Robinson & Beck, 2000, study 1). Others, especially those measuring children's experience of counterfactual emotions such as relief and regret, have found later development (e.g., Beck & Crilly, 2009; Guttentag & Ferrell, 2004, 2008; Rafetseder & Perner, 2012; Weisberg & Beck, 2010, 2012).

What counts as genuine, adult-like counterfactual thinking has been debated recently among developmental psychologists. Some researchers conceptualize counterfactual thinking broadly and argue that the ability to represent a premise that contradicts reality suffices as counterfactual thinking. This ability, which encompasses pretense, fiction, and future thinking, emerges early in development (Buchsbaum, Bridgers, Weisberg, & Gopnik, 2012; Weisberg & Gopnik, 2013). Beck and colleagues hold a more precise view of counterfactual thinking, contrasting real-world counterfactuals and general counterfactuals (the broad category encompassing fiction and pretend; Beck, 2016; Beck & Riggs, 2014). Real-world counterfactuals carry the additional requirements of bearing a close resemblance to reality and being held in comparison to reality in the mind of the reasoner. Beck and colleagues argue that the ability to reason about real-world counterfactuals develops around age 6. Rafetseder and colleagues provide a narrower still definition of counterfactual thinking, arguing that children do not display genuine counterfactual thinking until they appreciate what they call the nearest possible world constraint, which develops in adolescence (Rafetseder, Cristi-Vargas, & Perner, 2010; Rafetseder & Perner, 2014; Rafetseder, Schwitalla, & Perner 2013).

The nearest possible world constraint involves the manipulation of one aspect of an event while holding all other features constant. That is, one should change only those features that are casually dependent on a counterfactual antecedent and maintain all other features of the real world (Edgington, 2011). A good test of this is to present children with semifactual or causally overdetermined scenarios. Two independent antecedent events lead to the same outcome, and one is asked to consider how the outcome would be different if *one* of the antecedent events had not happened. For example, a character may get rained on and then go swimming, with both events leading to the outcome of "getting wet." Rafetseder et al. (2010, 2013) argue that children in Harris et al. (1996) study, who were presented with a character who dirties the floor with her muddy shoes, were passing using *basic conditional reasoning*, which simply requires children to reason using general regularities, such as "if no dirty shoes, then clean floor." In this case, the reasoner can disregard all other aspects of the scenario and just reason based on their causal knowledge about the impact of footwear cleanliness on floors.

To test the distinction between basic conditional and counterfactual reasoning, Rafetseder et al. (2010, 2013) presented children, adolescents, and adults with scenarios that required them to appreciate the nearest possible world constraint. One task required participants to learn a complex set of rules about a particular story. A second task presented simpler stories that did not require participants to learn a set of rules (Rafetseder et al., 2013, study 2). The simpler stories had the following format: "Susie and her brother Max are outside playing. They go into the kitchen and they don't take their shoes off. They make the floor all dirty." Participants were then asked a counterfactual question: "What would have happened if Susie had taken off her shoes? Would the floor be clean or would it be dirty?" Here, the correct answer is that the floor would still be dirty *even if* Susie had taken her shoes off because Max would still have his shoes on. It also was not until children were approximately 12 years old that they reliably passed on either set of tasks.

These results are quite surprising. We agree with Rafetseder and colleagues that appreciating the nearest possible world constraint is a valuable litmus test for *sophisticated* counterfactual thinking. However, we argue that certain features of Rafetseder et al. (2010, 2013) tasks may have underestimated children's appreciation of the nearest possible world constraint and therefore their counterfactual reasoning abilities.

First, the counterfactual questions asked were somewhat misleading from a pragmatic standpoint. An experimenter presents a short story with two similar events (e.g., two characters wearing muddy shoes) and then asks a question about the removal of only one of these events (e.g., "What would have happened if Susie had taken off her shoes?"). This question is free of certain markers one may typically use to communicate unambiguously (i.e., the experimenter does not say "*just* Susie" or "*only* Susie"). Participants may make the incorrect inference that the experimenter intended to refer to both events (e.g., Max and Susie both wearing their shoes). Another issue is that the very process of the experimenter asking the child a counterfactual question about such scenarios may lead the child to assume that the experimenter has implied a change from the status quo. If the experimenter and child have just watched a story together that ends with a dirty floor and the experimenter then asks the child about a change to an antecedent, the child may provide a "different" response, not because they are capable only of basic conditional reasoning but because they infer this is the answer the experimenter is looking for.

A second and overlapping issue is that participants may have made incorrect inferences about the causal structure of the events in some of the stories presented to them. Although Rafetseder et al. (2013) intended for the antecedent events to be causally unrelated to one another, children may have represented the events differently, perhaps inferring that the antecedents existed in a causal chain or both had a common cause. Removing one event in a counterfactual scenario may have had a knock-on effect, wiping out the other. Indeed, Rafetseder et al. (2013) reported that some children described a causal connection between the two events in their "sleeping" story. In the above example, some children may have inferred that Max would have taken his shoes off *because* Susie took her shoes off, given that their entry into the kitchen could be seen as a collaborative action in the original scenario. If this were the case, then "the floor would be clean" would actually be the correct answer—an answer Rafetseder et al. (2013) marked as incorrect. Thus, children who were generating inferences to increase coherence within the story could have been those most likely to fail the task. Children could have failed this task not by disobeying the nearest possible world constraint—which stipulates that only those features that are *causally dependent* on the antecedent should be changed—but instead by making incorrect judgments about what exactly was causally dependent on the antecedent. Because the causal structure children represented may have been different from the causal structure intended by the researchers, children's failure on counterfactual questions could have stemmed not from an inability to reason counterfactually but from different representations of reality off which their counterfactual inferences were based. In this study, we attempted to make the causal event structure in stories more transparent to ensure the *input* to children's counterfactual reasoning was as intended.

Rafetseder et al. (2010, 2013) findings present an apparent challenge for theories of causal reasoning that argue that counterfactual inferences are made either before (Harris et al., 1996; Lewis, 1973; Mackie, 1974) or along with causal inferences (Pearl, 2000; Schulz et al., 2007; Woodward, 2003). These theories suggest that children should be able to make counterfactual inferences that comply with their causal inferences. Consistent with this argument, Schulz et al. (2007) found that preschoolers were both able to use evidence from interventions to infer the correct causal structure of a simple physical system (a gear toy) and use their knowledge of causal structure to answer counterfactual questions about the outcome of hypothetical interventions to the toy. Preschoolers' causal and counterfactual inferences appeared to go hand in hand.

In contrast, a set of studies with older children (5- to 7-year-olds) found an asymmetry between children's causal and counterfactual judgments (Burns & McCormack, 2009; Frosch, McCormack, Lagnado, & Burns, 2012). Children's counterfactual judgments were not sensitive to differences in causal structure and were inconsistent with their causal judgments which they had to infer based on temporal cues. One possible reason for this discrepancy in findings may have to do with the quality of children's representations of causal structure and the way in which they are derived (McCormack, Frosch, & Burns, 2011). In particular, representations built from perceiving the temporal relation between events (as in Burns & McCormack, 2009; Frosch et al., 2012) may not support reasoning about counterfactual interventions.

The existing findings underscore the importance of correctly characterizing the representation children are considering in their counterfactual computations. Differences in children's representations of the causal structure of reality could lead to apparent failures in counterfactual reasoning. In the current research we examine this as a possible reason for children's failure at counterfactual reasoning tasks (Rafetseder et al., 2010, 2013). In particular, we investigated whether adapting Rafetseder et al. (2013, study 2) stories would allow children to demonstrate counterfactual reasoning by respecting the nearest possible world at a younger age. In Experiment 1, we adapted versions of Rafetseder et al.'s (2013) stories. In the new adapted versions, we attempted to make the causal structure of the events less ambiguous by making antecedent events causally unrelated to one another. We did this by, for example, separating the events temporally and by giving the protagonists different goals. In Experiment 2, we created four new sets of stories that were more tightly controlled and compared children's reasoning about stories with different causal structures—in which antecedent events were causally connected or causally disconnected. By presenting children with a clearer causal structure, we expected that they would show better performance at a younger age than children in the previous studies, as described earlier.

We chose to include 6-, 7-, and 8-year-olds in the present research—an age range during which we expected nearest possible world understanding to develop. Children in this age range did not succeed on Rafetseder et al.'s (2013) tasks, but we reasoned that, with the described modifications, children would demonstrate earlier understanding of the nearest possible world constraint. This age range also coincides with the age at which Beck and Riggs (2014) argue that adult-like counterfactual reasoning develops. In sum, we hypothesized that between the ages of 6 and 8 years, children would demonstrate nearest possible world understanding when the causal structure of the events was made less ambiguous.

Experiment 1

Method

Participants

Participants were recruited and tested at a science center in Toronto, Canada between January and April 2016. The final sample consisted of 98 children between the ages of 6.0 and 8.98 years (M = 7.44, SD = .88; 49 girls): thirty-four 6-year-olds (M = 6.48, SD = .33), thirty-two 7-year-olds (M = 7.42, SD = .30), and thirty-two 8-year-olds (M = 8.47, SD = .32). The sample size was determined using power analyses and to meet counterbalancing requirements. All participants were typically developing, and spoke and understood English (the language in which testing took place) fluently, according to parental report. For inclusion, children were required to speak and hear English 50% or more of the time. An equal number of girls and boys were tested across each age group and condition. An additional 18 children were tested but excluded because they did not meet language or age requirements (n = 13), answered both sets of control questions incorrectly (n = 2), or experimenter errors (n = 3). Signed parental consent and children's verbal assent were obtained for all participating children. Across both experiments, our sample of children were 54% Caucasian, 13% mixed ethnicity, 12% Chinese, 6% South Asian, 3% Japanese, 3% Latin American, 3% Southeast Asian, 3% West Asian, and 3% other ethnicity. The majority of children were from middle-class households.

Materials and Design

We adapted four stories from Rafetseder et al. (2013) study but made several changes to the stimuli and procedure, detailed next. As mentioned earlier, we hypothesized that children in Rafetseder and colleagues' study may have represented the events differently than intended by the researchers. To test this possibility, we created new versions of each of the four stories to make the causal structure of the events clearer to children. There were two versions of each story: an *ambiguous version*, based on Rafetseder and colleagues' stories, in which the causal relation between antecedents was unclear (e.g., children could have seen the events as related), and an *unrelated version*, in which we attempted to make the antecedent events causally independent. For example, in the muddy floor story, Susie and Max come inside and make the floor all dirty with their muddy shoes. In the *ambiguous condition*, Susie and Max come inside together. In the *unrelated condition*, there was a temporal separation between the two children entering the house, and the characters were given different motivations for entering the house (i.e., Susie comes inside to get some juice and Max later runs inside to get a bandage). The logic of the other three stories was the same. Unlike Rafetseder et al. (2013), who used props and enacted the scenarios live, we chose to present the stories to children using pictures and prerecorded audio stimuli in order to standardize the delivery of the stories. Visual stimuli for the eight stories were developed using the program Bitstrips Inc.: Toronto, Canada The story narration was prerecorded by a female native English speaker not otherwise involved in the study. Transcripts and images from all stories are available in the Supporting Information.

Procedure

Each participant was randomly assigned to one of two conditions (ambiguous or unrelated) and heard the four stories in one of two orders: (1) muddy floor, (2) sleeping, (3) getting wet, and (4) painting, or in a second randomized order (2, 1, 4, and 3). Children were presented with four prerecorded stories each followed by two control questions and one test question. The entire task took around 5 min. The procedure was the same for each age group and condition. Participants sat in front of a laptop and the session started with the experimenter introducing the stories by saying, "Today we're going to listen to some stories on the computer and then I'll ask you some questions about them. I want you to listen very carefully. Are you ready?" The story then played from start to finish and terminated with a black screen. While the black screen was displayed, the experimenter asked the child a series of questions. The experimenter first asked two control questions. The *before* control question asked for the initial state before the two antecedent events happened (e.g., "Was the floor dirty or clean before Susie and Max walked in?"), and the *now* control question asked for the final state (e.g., "Is the floor dirty or clean now?"). The experimenter then asked the *counterfactual test question* about either the first or second antecedent event (e.g., "<u>How would the story have ended</u> if Susie had taken her shoes off? Would the floor <u>still</u> be dirty, or would it be clean?"). The order of mention of the adjectives of interest (e.g., *dirty or clean* vs. *clean or dirty*) was counterbalanced. We changed the wording of the counterfactual test questions from Rafetseder et al.'s (2013) study (changes underlined) to make it clearer to the children what information the experimenter was requesting. First, we specified that the counterfactual requested information about a possible change to the state of affairs *at the end of the story*. Second, we inserted the word "still" into all counterfactual questions to make reference to the actual outcome.

Children were asked three questions after each story, in the following order: a before question, a now question, and a counterfactual question, as in previous studies using this design (Harris et al., 1996; Rafetseder et al., 2013). Note that for each story, children were asked a counterfactual question about *either* the first or second antecedent event. Each child was asked about the first antecedent (e.g., ". . . if Susie had taken her shoes off?") for two stories and the second antecedent (e.g., ". . . if Max had taken his shoes off?") for the other two stories. This was fully counterbalanced within and between participants.

Coding

The children's answers were live scored during testing. Sessions were also video recorded for reliability purposes. A second coder, unaware of the goal of the study, coded 30 out of the 98 included participants (31%). The coders had a high level of agreement (98%), $\kappa = .96$, $p < .001$.

Results and Discussion

Control Questions

Children were very accurate on the control questions about the state of affairs at the beginning and ending of the stories (i.e., the before and now questions). Overall, children were 96% accurate on before questions and 95% accurate on now questions. Children's responses to before questions were more accurate in the unrelated condition (98% correct) than in the ambiguous condition (93% correct), Mann–Whitney $U = 1{,}053.50$, $p = .047$. Children's responses to now questions did not differ significantly across conditions ($p = .648$). As mentioned earlier, children who answered both the before and now questions incorrectly were excluded from analyses ($n = 2$). Excluding those children who answered *any* control questions wrong from the analyses did not change the pattern of results, and therefore these children's data were retained in the analyses reported below.

Counterfactual Questions

Children received a score out of 4 based on their answers to the counterfactual questions across all four stories. Children's scores did not differ according to story order ($p = .624$). We used ordinal regression to investigate the effect of the two predictor variables, condition and age, on the overall score out of 4. Ordinal regression is a method for modeling ordinal outcome measures and yields odds ratio statistics that indicate the magnitude of the odds of a higher score given a change in the value of a predictor variable. Odds ratios provide an index of effect size. A model including condition (unrelated or ambiguous) as a categorical predictor, age (centered by subtracting the mean) as a continuous predictor variable, and the Condition × Age interaction was significant, $\chi^2(3) = 12.29$, $p = .006$. Children in the unrelated condition had 2.74 times higher odds of receiving a higher score than those in the ambiguous condition, Wald $\chi^2(1) = 7.39$, $p = .007$, parameter estimate = 1.009, 95% CI [0.28, 1.74]. Age did not emerge as a significant predictor ($p = .843$). The Condition × Age interaction was not significant but suggested a marginal effect of age on condition, odds ratio = 1.89, Wald $\chi^2(1) = 2.29$, $p = .130$, parameter estimate = 0.64, 95% CI [0.06, 1.33].

We also analyzed children's performance in each condition against chance (with chance = 2). Children in the unrelated condition answered significantly more questions correctly than expected by chance, Wilcoxon signed-rank test: $z = -2.58$, $p = .010$. Children in the ambiguous condition did not answer more questions correctly than expected by chance, $z = -1.36$, $p = .173$.

Questions About the First Versus Second Antecedent

The total score out of four comprised questions about the removal of the first antecedent event for two stories, and the second antecedent for the other two stories. Overall, a larger percentage of children answered both questions about the second antecedent (Q2: 40.8%) correctly than the first antecedent (Q1: 28.6%). This was also the case when looking within each condition (ambiguous Q1: 18.4% vs. Q2: 28.6%; unrelated Q1: 38.8% vs. Q2: 53.1%). To analyze whether this difference was statistically significant, we ran a generalized estimating equation (GEE) model with condition as a between-subjects predictor, question type (Q1 or Q2) as a within-subjects variable, and age (centered) as a covariate, with score (out of 2) as the dependent variable. This model had a multinomial distribution and a cumulative logit link function. All variables were entered simultaneously in the model. We did not find evidence for a statistically significant effect of question type on score, $B = -0.38$, $SE = .29$, Wald $\chi^2(1) = 1.77$, $p = .183$, odds ratio = 0.68, 95% CI [0.39, 1.20]. Consistent with the results of the ordinal regression presented above, condition was a significant predictor, $B = 1.12$, $SE = .39$, Wald $\chi^2(1) = 8.36$, $p = .004$, odds ratio = 3.07, 95% CI [1.44, 6.58], whereas age was not, odds ratio = 1.35, $p = .097$. The Condition × Question Type interaction was not significant ($p = .687$). Figure 1 displays the percentage of children receiving scores of 0, 1, and 2 on each question type in each condition.

In sum, children in this study performed well when presented with stories in which the two antecedent causes were separated from one another. Recall that children in Rafetseder et al.'s (2013) study did not answer counterfactual questions correctly until age 12. We adapted Rafetseder et al.'s stories in order to clarify the causal structure of the events. Children performed significantly better in this new causally unrelated condition both compared to chance and to the ambiguous stories modeled closely after Rafetseder et al.'s (2013) vignettes. Children in our study were able to obey the nearest possible world constraint substantially earlier than children in Rafetseder et al.'s (2013) study. We suggest that children's weaker performance in this previous study was due in part to features of the tasks used, rather than limitations in children's counterfactual reasoning.

In the attempt to make the antecedent events causally unrelated in the current study, while maintaining the main features of Rafetseder et al. (2013) examples for comparison purposes, it was necessary to add additional context (e.g., giving Susie and Max different goals for entering the kitchen). A concern therefore is that the better performance in this condition may have been due to the additional context, rather than removing the possibility of a causal relation between antecedents. Another concern is that some of the stories in the unrelated condition left open the possibility that antecedent events were still related to one another, and therefore, the causal structure of these stories may not have been as clear as intended. For example, in the muddy floor story the floor was clean when Max entered in the ambiguous version but already dirty when he entered in the unrelated version. In the latter case, children may have inferred that he would remove or keep his shoes on depending on the state of the floor when he entered the room. Although we had Max rush in to the kitchen in the unrelated version, children may still have represented a different causal structure than intended.

The fact that children performed slightly better when asked about the removal of the second antecedent than the first in both conditions suggests that they may have been drawing causal inferences even in the unrelated condition. An alternate

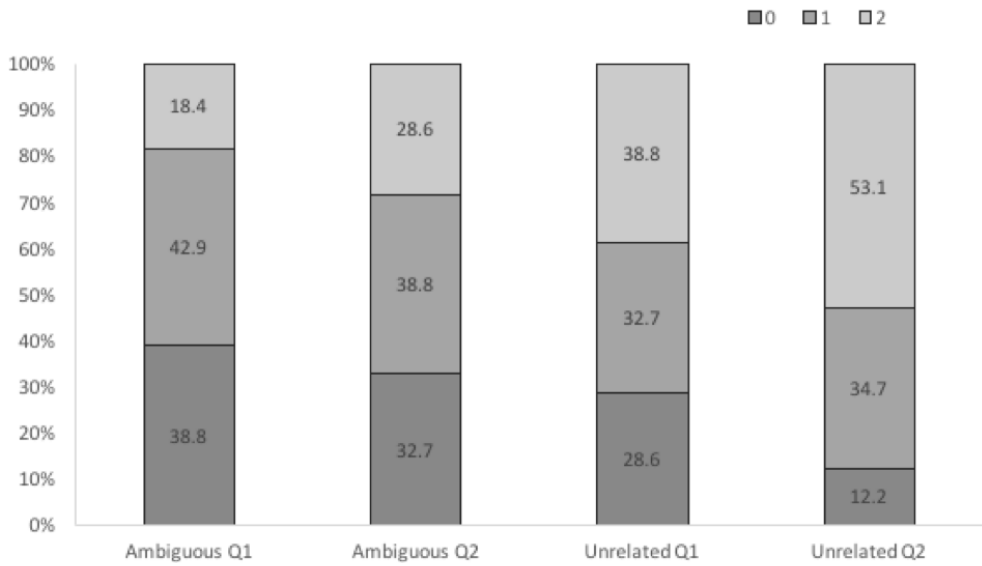

Figure 1. Percentage of children in each condition scoring at each level out of 2 for questions about the first antecedent (Q1) and second antecedent (Q2) in Experiment 1.

explanation for these findings is that children may have found it easier to access or construct a representation of the state of affairs in the absence of the second event. By virtue of the temporal sequence, children saw the scene after the first antecedent event occurred, but before the second had occurred, and therefore may have found it easier to answer the question about the removal of the second antecedent. This may particularly have been the case in the unrelated condition, in which the two antecedent events had more of a temporal separation between them than in the ambiguous condition.

In Experiment 2, we created four new sets of stories that were more tightly controlled and that had a clearer causal structure.

Experiment 2

We created new stories to equate the two versions of each story and to test the generalizability of the findings from Experiment 1. The stories followed the same structure as those in Experiment 1, with two antecedent events leading to the same outcome. However, in one order, the same two antecedent events were causally connected to one another and in the reversed order were causally disconnected. We included a continuous age range in Experiment 1 (from 6.0 to 8.98 years) but decided to include only 6- and 8-year-olds in the present experiment and treated age as a categorical variable, because children's performance did not appear to change drastically between the ages of 6 and 7 in Experiment 1. We hypothesized that by age 8 children would be able to negotiate these different causal structures in their counterfactual representations and would perform above chance in both conditions. We included questions about the removal of both the first and second antecedent in this experiment but were not necessarily expecting a difference in performance across question types. If children were correctly representing the causal structure, one would not expect to see a difference across question types.

Method

Participants

The final sample consisted of 64 children in two age groups including thirty-two 6-year-olds (M = 6.70, SD = .25) and thirty-two 8-year-olds (M = 8.50, SD = .31). Language requirements were the same as in Experiment 1, and an equal number of girls and boys were tested across age groups and conditions. Three additional children were tested and had to be excluded because of failing the control questions (n = 1) and insufficient English language exposure (n = 2). Parental consent and child assent were obtained for all participating children. Children were tested in a science center (n = 25) or in our university laboratory (n = 38), drawn from a large database of area families in Toronto, Canada between April and December 2016. Performance did not differ significantly between children tested in the two locations, $p = .476$.

Materials

We developed two new sets of four stories. All stories consisted of two antecedent events that both lead to the same outcome. In the *causally connected* condition, the first antecedent was a cause of the second antecedent, and both led to the same outcome. In the *causally disconnected* condition, the first and second antecedents were not causally related to one another, and both led to the same outcome. Crucially, both versions contained the exact same antecedent events. In one order, they were causally connected, and in the reverse order, they were not. For example, in the connected version of the painting story, Heidi notices that one of the paints is in a leaky cup, and when she tries to pick up the cup to stop it from leaking, she knocks another paint cup over. In the disconnected version, the two events were in reverse order, such that Heidi first accidentally knocks over one of her paints before noticing the other cup leaking. Each story consisted of three pictures designed in Bitstrips Inc. and the story narration was prerecorded by the same speaker as in Experiment 1. Transcripts and images from each story are available in the Supporting Information.

Design and Procedure

Participants were assigned to one of the two conditions (connected or disconnected) and heard the stories in one of two orders (1) painting, (2) flowers, (3) dress, and (4) sandcastle or the reversed order (4, 3, 2, and 1). Participants were asked *before* and *now* control questions, and a *counterfactual* test question (e.g., "How would the story have ended if Heidi hadn't knocked over the cup of paint (the paint cup hadn't been leaking)?"), as in Experiment 1. Again, participants were asked questions about the first antecedent for two stories and the second antecedent for the other two stories. The correct answer was the same for the first and second antecedent question in the disconnected condition and the second antecedent question in the connected condition. In these cases, the correct answer was always that the outcome would still be the same (e.g., the table would still be covered in paint) because one of the two antecedents would have still occurred. The correct answer was different, however, when asked about the removal of the first antecedent in the connected condition. In this case, the outcome would be different (e.g., the table would be clean), because the first antecedent caused the second antecedent. A feature of the connected condition therefore is that children could arrive at the correct answer half of the time by using a simpler reasoning strategy—what Rafetseder and colleagues (2010, 2013; Perner & Rafetseder, 2011; Rafetseder & Perner, 2014) refer to as basic conditional reasoning. On the question about the first antecedent in the connected condition, both basic conditional and mature counterfactual reasoning give the same answer. This is an issue we will return to in the discussion in Experiment 2.

Coding

As in Experiment 1, live coding was used for all participants. A second coder, blind to the purpose of the study, scored the video recordings of 48 (75%) of the 64 included participants. Cohen's kappa was run and reached a good level of agreement (87%), $\kappa = .67$, $p < .001$. A third coder checked discrepancies (13% of trials), and in all cases they were due to poor audibility in the video recordings. Thus, we included the codes from the live coder in analyses.

Results and Discussion

Control Questions

Children's responses on the control questions were very accurate for both before (91%) questions and now questions (99%). Accuracy did not differ significantly between the two conditions for either before or now questions, $p = .104$ and $.557$, respectively.

Counterfactual Questions

As in Experiment 1, children received a score out of 4 based on their answers to the counterfactual questions across the four stories. Children who received the stories in the second order received marginally higher scores than those who received the stories in the first order ($p = .104$). We conducted an ordinal regression with counterfactual score as the dependent measure, and age (6 or 8 years) and condition (connected or disconnected) as categorical predictors, and the Condition × Age interaction term. This model was significant, $\chi^2(3) = 20.23$, $p < .001$. Eight-year-olds had 17.34 times higher odds of receiving a higher score than

6-year-olds, Wald $\chi^2(1) = 14.76$, $p < .001$, parameter estimate = −2.85, 95% CI [−4.31, −1.40]. Being in the connected versus disconnected condition was not associated with significantly different odds of receiving a higher score, odds ratio = 1.08, $p = .912$. The Age × Condition interaction was significant, odds ratio = 7.74, $\chi^2(1) = 4.38$, $p = .036$, parameter estimate = 2.05, 95% CI [0.13, 3.96]. It was only in the disconnected condition that 8-year-olds had higher odds of receiving a higher score than 6-year-olds, odds ratio = 13.11, Wald $\chi^2(1) = 11.01$, $p = .001$, parameter estimate = −2.57, 95% CI [−4.09, −1.05]. Eight-year-olds were not significantly more likely to receive a higher score than 6-year-olds in the connected condition, odds ratio = 2.30, $p = .216$.

We also compared children's performance against chance using a Wilcoxon signed-rank test with chance equal to 2. A Bonferroni correction was applied for multiple comparisons, with an adjusted alpha value of .008. Overall, children in the connected condition answered significantly more questions correctly than expected by chance, $z = -4.22$, $p < .001$. Children in the disconnected condition did not answer more questions correctly than expected by chance, $z = -1.68$, $p = .094$. Eight-year-olds answered significantly more questions correctly than expected by chance in both the connected, $z = -3.22$, $p = .001$, and disconnected conditions, $z = -2.92$, $p = .003$. Six-year-olds answered significantly more questions correctly than expected by chance in the connected condition, $z = -2.76$, $p = .006$, but not in the disconnected condition, $p = .294$.

To further investigate the fact that 6-year-olds' performance was better in the connected than the disconnected condition, we looked at their answers to questions about the first or second antecedents separately across conditions. Recall that children in the connected condition could have arrived at the correct answer to questions about the first antecedent via basic conditional reasoning. Six-year-olds answered more questions about the first antecedent correctly in the connected condition (Mdn = 2) than in the disconnected condition (Mdn = 1), Mann–Whitney $U = 68.00$, $z = -2.47$, $p = .014$. They answered marginally more questions correctly about the second antecedent in the connected condition (Mdn = 2) than in the disconnected condition (Mdn = 1), $p = .062$. Moreover, 6-year-olds' performance did not differ significantly across questions about the first versus second antecedent within either the connected, $p = .713$, or disconnected condition, $p = 1.00$.

Questions About the First Versus Second Antecedent

In the connected condition, a similar percentage of children answered both questions correctly about the first (Q1: 71.9%) and second antecedent (Q2: 68.8%). Within the disconnected condition, a similar percentage of children also answered both questions correctly about the first (Q1: 46.9%) and second antecedents (Q2: 53.1%). To analyze Q1 versus Q2 scores, we again conducted a GEE model with condition and age as between-subjects predictors, and question type (Q1 or Q2) as a within-subjects variable, and score (of 2) as the dependent variable. This model had a multinomial distribution and a cumulative logit link function. All variables were entered simultaneously in the model. Question type was not a significant predictor of score, $B = -0.68$, $SE = .77$, Wald $\chi^2(1) = 0.78$, $p = .376$, odds ratio = 0.51, 95% CI [0.11, 2.29]. Only age emerged as a significant predictor of score in this model, $B = -2.45$, $SE = .80$, Wald $\chi^2(1) = 9.30$, $p = .002$, odds ratio = 0.09, 95% CI [0.02, 0.42]. All other main effects and interactions were nonsignificant, $p = .270$ to 1. Figure 2 displays the percentage of children receiving scores of 0, 1, and 2 on each question type in each condition.

Eight-year-olds in this experiment made accurate counterfactual judgments both when faced with causally connected and causally disconnected antecedent events, as was the case in the unrelated condition in Experiment 1. Six-year-olds performed better in the connected condition than the disconnected condition. There are a few possible explanations for this finding. This better performance in the connected condition could have been partially due to the fact that children could succeed half of the time by using basic conditional reasoning. In support of this possibility, children were more accurate on questions about the first antecedent in the connected condition, where basic conditional and counterfactual reasoning would give the same response than in the disconnected condition, in which the two would give different responses. However, this cannot be the sole explanation as shown by the lack of significant differences between children's responses to questions about the first versus second antecedent within the connected condition and their marginally better performance on questions about the second antecedent in the connected versus disconnected condition.

A different explanation is that the increased causal coherence of the stories in the connected condition contributed to children's success. Decades of research have demonstrated that children's

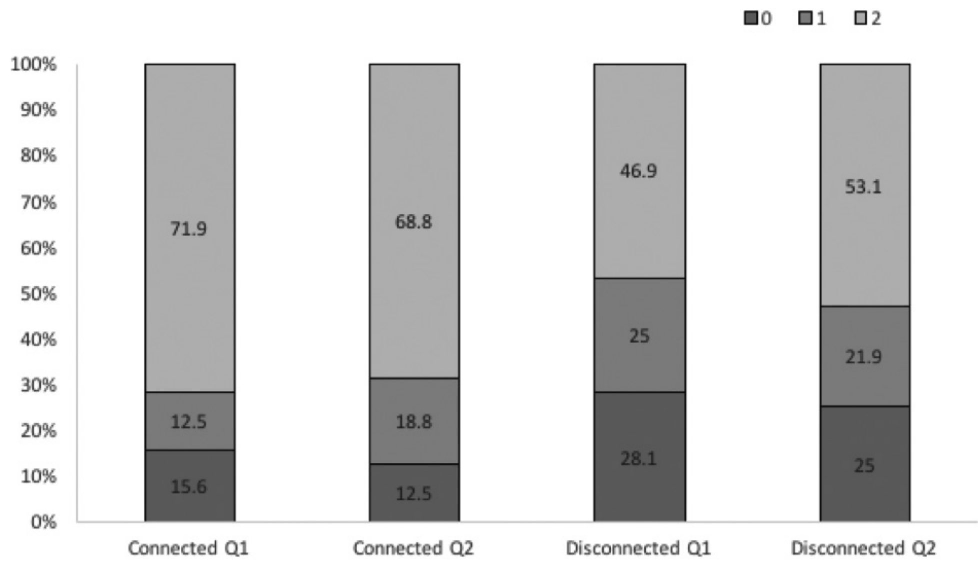

Figure 2. Percentage of children in each condition scoring at each level out of 2 for questions about the first antecedent (Q1) and second antecedent (Q2) in Experiment 2.

comprehension and memory is better for narratives with higher levels of causal coherence (e.g., Van Den Broek, Lorch, & Thurlow, 1996; Van Den Broek et al., 2005). Six-year-olds may therefore benefit from scenarios that are easier to comprehend and recall. Future work with this age group should focus on presenting simpler tasks that may not place as many demands on their comprehension (e.g., inferencing) and memory abilities.

The fact that 6-year-olds performed better in the connected than the disconnected condition lends support to the suggestion that children may seemingly fail on tasks of counterfactual reasoning because they make additional and unexpected causal inferences. In the connected condition, in which causal connections were already in place, there was little room for additional inferences. In the disconnected condition, although intended to present independent antecedent events, one could still potentially generate an explanation for how the events could be connected. Children could have drawn forward causal inferences or backtracking inferences, if they assumed that certain events were a foregone conclusion.

General Discussion

Mature counterfactual reasoning involves the ability to exercise the nearest possible world constraint. That is, the reasoner should change only those features of reality that are causally dependent upon the antecedent and hold all else constant (Edgington, 2011). Prior research found that children did not reason according to this constraint until age 12 (Rafetseder et al., 2010, 2013). The current results show that this ability is in place by age 8. Eight-year-olds in our study were able to correctly answer counterfactual questions both when antecedent events were causally connected to one another and when they were disconnected. Children's performance in this research indicates that they are able to contend with different causal structures in their counterfactual representations.

Why did children succeed much earlier on our version of the tasks requiring them to consider overdetermined scenarios than on Rafetseder et al.'s (2013) version? Our findings underscore the importance of taking task demands into account as well as the possibility that children may interpret events in a way not intended by the researchers. As we suggested in the Introduction, children may have performed poorly in Rafetseder et al.'s (2013) study because their representation of reality conformed to a different causal structure than intended by the researchers.

Limited previous research has found evidence for children's sensitivity to causal structure when reasoning counterfactually. Using simple physical systems, Schulz et al. (2007) found that preschoolers were able to make correct counterfactual inferences that respected different casual structures. In contrast, Burns and colleagues (Burns & McCormack, 2009; Frosch et al., 2012) found that 7-year-olds' counterfactual inferences did not respect the causal structure of a simple physical system, though this could have been because they had insufficient representations of the causal structure (McCormack et al., 2011). Although the children in our study succeeded at a later age than those in Schulz et al.'s (2007) study, together these findings suggest that children can reason counterfactually about events with a range of causal structures if the causal structure of the events is clear to them and their representation is robust. Unlike the Schulz et al. (2007) studies, ours used short stories involving agents and focused on older children's reasoning about causally overdetermined outcomes, which are arguably more challenging and require children to exercise the nearest possible world constraint.

We found striking preliminary evidence that children as young as 6 may be able to reason counterfactually about overdetermined events. Six-year-olds performed well in the connected condition in Experiment 2, which we suggested may have been due to increased sensitivity to the causal coherence of stories. We consider two additional explanations for the finding that younger children in the current experiments struggled under certain condition: difficulties with pragmatic inferences and difficulties creating separate event representations.

Inherent to the type of task used in this research are some features that make it pragmatically challenging. These were not issues we could escape. We made minor changes to wording compared to Rafetseder et al. (2013), but some of the more major issues, such as the fact that the experimenter refers to only one of two events, are fundamental features of the task. One could imagine asking the child "what if Susie had taken her shoes off, but Max had still left them on?," but would not be getting away from the issues that Rafetseder and colleagues laid out about the contrast between basic conditional and counterfactual reasoning. With minor pragmatic adjustments, children's performance was at chance in the younger age group in most of our conditions. These younger children still could have made incorrect pragmatic inferences, assuming perhaps that the experimenter intended to refer to both Susie *and* Max even when she only mentioned Susie. As children get older, they may become more accustomed to the fact that people are not always helpful in communication and as a result may perform better on tasks of this type. Future work should use simpler tasks that reduce the possibility of children making additional inferences. Tasks employing stories seem to be especially susceptible to these problems, and future work may make use of simple physical causation tasks or tasks involving live agents.

Events that can be confused pragmatically are also likely to be those that can be conflated conceptually. For example, children may construct a single event representation of "wearing dirty shoes" because two events involving characters wearing their muddy shoes inside are conceptually similar to one another. Thus, another possible explanation for failure on tasks similar to those used in the present study is that children construct a single event representation for similar events. This single event is entered into the counterfactual computation and leads to an incorrect response. In future studies, we plan to present children with scenarios in which two events are conceptually very distinct (e.g., a tree gets sick because someone peels its bark and because it gets struck by lightning) and events that are separated temporally and spatially, to ensure children are creating separate event representations in tasks that require them to exercise the nearest possible world constraint.

It is also worth mentioning that one study by Rafetseder and Perner (2010) found that children were able to engage in counterfactual reasoning by the age of 6. They gave children change-of-location counterfactual scenarios, in which a character was in a typical or atypical location when he was called to another location. For example, a doctor was either at the hospital (typical location) or at the park (atypical location) when he was called to the swimming pool for an emergency. When asked where the doctor would be if he had not been called, 6-year-olds demonstrated counterfactual reasoning. Importantly, this task seems to be free of many of the demands we have outlined above that are features of overdetermined scenarios. This task appears to be more pragmatically transparent, does not include easily conflated events, and does not license additional causal inferences. These features may account for children's earlier success compared to Rafetseder et al.'s (2010, 2013) studies.

The current findings raise the possibility that children may not respond in adult-like ways on counterfactual tasks not because they cannot reason counterfactually but because they approach these tasks differently, perhaps imputing additional causal connections that adults would view as unwarranted. With age and experience, individuals may become more conservative about the types of causal inferences they draw. Consistent with this argument, Lucas, Bridgers, Griffiths, and Gopnik (2014) found that children were more flexible than adults in the inferences they drew about novel causal systems than adults. These differences in causal inferencing could mean that, given the same events, adults and children enter different causal models into their counterfactual computations, resulting in diverging counterfactual inferences.

Conclusion

Many previous studies of the development of counterfactual reasoning have found successful performance by around the age of 4 (German & Nichols, 2003; Guajardo, Parker, & Turley-Ames, 2009; Guajardo & Turley-Ames, 2004; Harris et al., 1996; Riggs et al., 1998). These studies primarily required children to reason about counterfactual conditionals from counterfactual antecedents, without having to incorporate other aspects of reality into their counterfactual representation (Rafetseder & Perner, 2014) and without having to consider multiple possibilities simultaneously (Beck & Riggs, 2014). Studies that have required children to consider multiple possibilities (e.g., reality and a counterfactual possibility) have found that children do not succeed until age 5 or 6 (Beck & Guthrie, 2011; Beck et al., 2006; Perner et al., 2004; Rafetseder & Perner, 2010). Those that have required children to exercise the nearest possible world constraint have found that children do not pass until as late as adolescence (Rafetseder et al., 2010, 2013). However, the current findings demonstrate that given scenarios with clearer causal structures, children show success using the nearest possible world constraint between the ages of 6 and 8.

The current results suggest that important developments in children's counterfactual reasoning take place between the ages of 3.5, when children passed Harris et al. (1996) counterfactual conditional task, and 8, when children showed robust performance on our tasks. Further developments may also take place between the ages of 8, when children succeed on our task, and 12, when children succeeded on Rafetseder et al. (2010, 2013) tasks. In particular, these findings raise questions about the nature of the development that takes place—whether it is primarily in counterfactual reasoning or related abilities such as causal reasoning, pragmatic inferencing, or executive functioning. Although the current results indicate that children are able to reason about overdetermined events with different causal structures by age 8, future work should investigate the range of causal models children can manipulate counterfactually, and how this may be related to their domain-specific knowledge and the (ir)relevant causal inferences they make.

References

Beck, S. R. (2016). Counterfactuals matter: A reply to Weisberg & Gopnik. *Cognitive Science*, *40*, 260–261. https://doi.org/10.1111/cogs.12242

Beck, S. R., & Crilly, M. (2009). Is understanding regret dependent on developments in counterfactual thinking? *British Journal of Developmental Psychology*, *27*, 505–510. https://doi.org/10.1348/026151008X401697

Beck, S. R., & Guthrie, C. (2011). Almost thinking counterfactually: Children's understanding of close counterfactuals. *Child Development*, *82*, 1189–1198. https://doi.org/10.1111/j.1467-8624.2011.01590.x

Beck, S. R., & Riggs, K. J. (2014). Developing thoughts about what might have been. *Child Development Perspectives*, *8*, 175–179. https://doi.org/10.1111/cdep.12082

Beck, S. R., Riggs, K. J., & Gorniak, S. L. (2010). The effect of causal chain length on counterfactual conditional reasoning. *British Journal of Developmental Psychology*, *28*, 505–521. https://doi.org/10.1348/026151009X450836

Beck, S. R., Robinson, E. J., Carroll, 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

Burns, P., & McCormack, T. (2009). Temporal information and children's and adults' causal inferences. *Thinking & Reasoning*, *15*, 167–196. https://doi.org/10.1080/13546780902743609

Byrne, R. M. J. (2016). Counterfactual thought. *Annual Review of Psychology*, *67*, 7.1–7.23. https://doi.org/10.1146/annurev-psych-122414-033249

Edgington, D. (2011). Causation first: Why causation is prior to counterfactuals. In C. Hoerl, T. McCormack, & S. R. Beck (Eds.), *Understanding counterfactuals, understanding causation: Issues in philosophy and psychology* (pp. 230–241). Oxford, UK: Oxford University Press.

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

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

German, T. P., & Nichols, S. (2003). Children's counterfactual inferences about long and short causal chains. *Developmental Science*, *6*, 514–523. https://doi.org/10.1111/1467-7687.00309

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

Guajardo, N. R., Parker, J., & Turley-Ames, K. (2009). Associations among false belief understanding, counterfactual reasoning, and executive function. *British Journal* of Developmental Psychology, 27, 681–702. https://doi.org/10.1348/026151008X357886

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

Guttentag, R., & Ferrell, J. (2004). Reality compared with its alternatives: Age differences in judgments of regret and relief. *Developmental Psychology*, *40*, 764. https://doi.org/10.1037/0012-1649.40.5.764

Guttentag, R., & Ferrell, J. (2008). Children's understanding of anticipatory regret and disappointment. *Cognition & Emotion*, *22*, 815–832. https://doi.org/10.1080/02699930701541542

Harris, P., 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

Lewis, D. (1973). *Counterfactuals*. Oxford, UK: Basil Blackwell.

Lucas, C. G., Bridgers, S., Griffiths, T. L., & Gopnik, A. (2014). When children are better (or at least more open-minded) learners than adults: Developmental differences in learning the forms of causal relationships. *Cognition*, *131*, 284–299. https://doi.org/10.1016/j.cognition.2013.12.010

Mackie, J. L. (1974). *The cement of the universe: A study of causation*. Oxford, UK: Oxford University Press.

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.

Pearl, J. (2000). *Causality: Models, reasoning and inference*. New York, NY: Cambridge University Press.

Perner, J., & Rafetseder, E. (2011). Counterfactual and other forms of conditional reasoning: Children lost in the nearest possible world. In C. Hoerl, T. McCormack, & S. R. Beck (Eds.), *Understanding counterfactuals, understanding causation: Issues in philosophy and psychology* (pp. 99–109). Oxford, UK: Oxford University Press.

Perner, J., Sprung, M., & Steinkogler, B. (2004). Counterfactual conditionals and false belief: A developmental dissociation. *Cognitive Development*, *19*, 179–201. https://doi.org/10.1016/j.cogdev.2003.12.001

Rafetseder, E., Cristi-Vargas, R., & Perner, J. (2010). Counterfactual reasoning: Developing a sense of "nearest possible world." *Child Development*, *81*, 376–389. https://doi.org/10.1111/j.1467-8624.2009.01401.x

Rafetseder, E., & Perner, J. (2010). Is reasoning from counterfactual antecedents evidence for counterfactual reasoning? *Thinking & Reasoning*, *16*, 131–155. https://doi.org/10.1080/13546783.2010.488074

Rafetseder, E., & Perner, J. (2012). When the alternative would have been better: Counterfactual reasoning and the emergence of regret. *Cognition & Emotion*, *26*, 800–819. https://doi.org/10.1080/02699931.2011.619744

Rafetseder, E., & Perner, J. (2014). Counterfactual reasoning: Sharpening conceptual distinctions in developmental studies. *Child Development Perspectives*, *8*, 54–58. https://doi.org/10.1111/cdep.12061

Rafetseder, E., Schwitalla, M., & Perner, J. (2013). Counterfactual reasoning: From childhood to adulthood. *Journal of Experimental Child Psychology*, *114*, 389–404. https://doi.org/10.1016/j.jecp.2012.10.010

Riggs, K. J., Peterson, D. M., Robinson, E. J., & Mitchell, P. (1998). Are errors in false belief tasks symptomatic of a broader difficulty with counterfactuality? *Cognitive Development*, *13*, 73–90. https://doi.org/10.1016/S0885-2014(98)90021-1

Robinson, E. J., & Beck, S. R. (2000). What is difficult about counterfactual reasoning? In P. Mitchell & K. Riggs (Eds.), *Children's reasoning and the mind* (pp. 101–120). Hove, UK: Psychology Press.

Schulz, L. E., Gopnik, A., & Glymour, C. (2007). Preschool children learn about causal structure from conditional interventions. *Developmental Science*, *10*, 322–332. https://doi.org/10.1111/j.1467-7687.2007.00587.x

Sobel, D. M. (2011). Domain-specific 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.

Van Den Broek, P., Kendeou, P. A., Kremer, K., Lynch, J., Butler, J., White, M. J., & Lorch, E. P. (2005). Assessment of comprehension abilities in young children. In S. Stahl & S. Paris (Eds.), *Children's reading comprehension assessment* (pp. 107–130). Mahwah, NJ: Erlbaum.

Van Den Broek, P., Lorch, E. P., & Thurlow, R. (1996). Children's and adults' memory for television stories: The role of causal factors, story-grammar categories, and hierarchical level. *Child Development*, *67*, 3010–3028. https://doi.org/10.2307/1131764

Weisberg, D. P., & Beck, S. R. (2010). Children's thinking about their own and others' regret and relief. *Journal of Experimental Child Psychology*, *106*, 184–191. https://doi.org/10.1016/j.jecp.2010.02.005

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

Weisberg, D. S., & Gopnik, A. (2013). Pretense, counterfactuals, and Bayesian causal models: Why what is not real really matters. *Cognitive Science*, *37*, 1368–1381. https://doi.org/10.1111/cogs.12069

Woodward, J. (2003). *Making things happen: A theory of causal explanation*. New York, NY: Oxford University Press.

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's website:

Data S1. Experiment 1 and Experiment 2 Stimuli