DOI: 10.1111/cdev.13931

EMPIRICAL ARTICLE

CHILD DEVELOPMENT

Children's hypothetical reasoning about complex and dynamic systems

Angela Nyhout[1,2] | **Hilary Sweatman**[3] | **Patricia A. Ganea**[2]

[1]School of Psychology, University of Kent, Canterbury, UK
[2]Department of Applied Psychology & Human Development, University of Toronto, Toronto, Ontario, Canada
[3]Department of Neurology and Neurosurgery, McGill University, Montreal, Québec, Canada

Correspondence
Angela Nyhout, School of Psychology, University of Kent, Canterbury CT2 7FS, UK.
Email: a.nyhout@kent.ac.uk

Funding information
Natural Sciences and Engineering Research Council of Canada, Grant/Award Number: RSD1270194; Social Sciences and Humanities Research Council of Canada, Grant/Award Number: 430-2019-01128; University of Toronto

Abstract
Children's hypothetical reasoning about a complex and dynamic causal system was investigated. Predominantly White, middle-class 5- to 7-year-old children from the Greater Toronto Area learned about novel food chains and were asked to consider the effects of removing one species on the others. In Study 1 ($N=72$; 36 females, 36 males; 2018), 7-year-olds answered questions about both direct and indirect effects with a high degree of accuracy, whereas 5-year-olds performed at chance. Six-year-olds showed intermediate performance. Using food chains with clearer constraints, Study 2 ($N=72$; 35 females, 37 males; 2020–2021) replicated these findings. These results indicate that the ability to think about hypothetical changes to dynamic causal systems develops between 5 and 7 years. Implications for science education are discussed.

A number of books and documentaries have recently invited their readers and viewers to entertain a frightening premise: a world without bees. In this hypothetical world, no bees exist to pollinate crops, hundreds of varieties of crops are lost, supermarket shelves sit empty, and humans and many other species that depend on these crops begin to die off. Imagining alternative representations of the world may help individuals to understand the causal role these insects play in our ecosystem and may compel behavioral change (e.g., growing bee-friendly plants). Hypothetical thought experiments like this one involve simulating a plausible future. Others may be more implausible. For instance, in Stephen Jay Gould's "replaying life's tape" thought experiment, he asks whether the natural world would still exist as we know it if we were able to rerun evolution (Gould, 1990). The ability to consider hypothetical pasts (counterfactual thinking) and futures (future hypothetical thinking) enables the reasoner to make more optimal decisions and predictions, allowing, for instance, for the effects of species loss—as in the example of bees—to be predicted and mitigated before it is too late. Thought experiments are a common tool in science education (Gilbert & Reiner, 2000) and may support the learning of complex concepts such as evolution, global warming, and forces and motion.

Our interest in the current study was in a specific type of thought experiment in which the learner imagines hypothetical changes to a complex and dynamic causal system, common in the biological domain. We consider cases in which the reasoner takes a representation of a causal system—in this case, a food chain—and imagines changes to it, while holding constant causal relations in the system to draw inferences about the imagined change. Recently, Nyhout and Ganea (2019a) argued that this is the same process we follow when reasoning about episodic counterfactuals: the reasoner first takes a representation of reality—a specific past event in the case of episodic counterfactual thinking, or a causal

Abbreviation: GEE, generalized estimating equation model in the present case—then introduces a hypothetical change to it and reasons through the implications of this change. Whereas most past developmental work has dealt with episodic counterfactual thinking involving agents in a specific event, previous work has not, to our knowledge, investigated children's ability to reason hypothetically about generic causes that are often of interest in science learning—termed *type* causation (see also De Brigard & Parikh, 2019 for the related notion of *semantic* counterfactuals). At this stage, it is important to establish whether children can engage in this type of complex reasoning before investigating whether and how it may contribute to learning. Most past developmental research has focused almost exclusively on what philosophers have termed *token* causation involving singular events, with a particular focus on children's reasoning about agents in simple cause–effect relations in narratives (examples for key terms are provided in Table 1).

Children's counterfactual reasoning about such token causes shows multiple and protracted developments (e.g., Beck & Riggs, 2014; Rafetseder & Perner, 2014), with many aspects in place in the preschool years (e.g., Buchsbaum et al., 2012; Harris et al., 1996; Nyhout & Ganea, 2019b) and others not reaching maturity until middle to late childhood (e.g., Beck & Crilly, 2009; O'Connor et al., 2012; Rafetseder & Perner, 2012). This picture of the developmental profile as protracted is partially attributable to differences in the complexity of tasks children are asked to perform, with children showing later success when they are asked to reason about more complex causal relations (Nyhout & Ganea, 2019a). But it also has to do with disagreements between researchers over how to conceptualize counterfactual and hypothetical thinking. Whereas some researchers argue that counterfactual reasoning encompasses a broad swath of hypothetical thinking abilities—including pretend play, future hypotheticals, and past counterfactuals—others argue there is a fundamental distinction between counterfactuals about past episodes and other forms of hypothetical thinking (for a debate, see the exchange between Beck, 2016; Weisberg & Gopnik, 2013, 2016). Other researchers argue that children should only be credited with the ability to think counterfactually when they obey what they call the *nearest possible world constraint*: the reasoner should change only those features of a situation that are dependent on a counterfactual premise, while holding all else constant (Edgington, 2011; Perner & Rafetseder, 2011; Rafetseder et al., 2010). Nyhout and Ganea (2019a) argue that counterfactual reasoning should be characterized by its underlying cognitive processes. That is, taking a *representation*, introducing a change to it, and reasoning about the implications of the change—but apply this to a broader range of thinking involving both specific episodes (episodic counterfactual thinking) and semantic or generic causal systems (hypothetical thinking) that share a common underlying reasoning process.

In the current study, we use Nyhout and Ganea's (2019a) framework and extend our investigation to an as-yet unexplored aspect of hypothetical thinking. We refer to the reasoning under investigation in the current study as *hypothetical thinking*, because the problems were not specifically focused on past events and children could answer with reference to hypothetical future changes, as is typically the case with type causation (Woodward, 2011).

Overall, existing developmental findings tell us little about how children reason hypothetically about complex causal systems and different types of causation. Whereas most studies have investigated children's reasoning about simple cause–effect relations involving *agents* in narratives, only a few studies have looked at children's counterfactual and conditional reasoning about *physical causal systems* (Frosch et al., 2012; McCormack et al., 2009, 2016, 2018; Nyhout & Ganea, 2019b). Even in these studies, the events in question were *single occurrences* (i.e., token causation) as in the previous studies involving questions about agents, and the physical apparatuses (e.g., ramps and machines) were operated by human agents. Limited work has also looked beyond simple cause–effect relations to investigate children's reasoning about *causal chains* involving agents in narratives. In everyday reasoning, we often set forth a counterfactual premise and make inferences not only about the implications for a cause's direct outcome but also how it may change other variables indirectly. German and Nichols (2003) found that children were better able to reason counterfactually about direct outcomes in short causal chains than indirect outcomes in longer chains, because reasoning about long causal chains requires extra inferential steps and therefore places more demands on executive functions and information processing abilities. However, this effect of differences in

TABLE 1 Key terms.

Term	Definition	Example
Hypothetical thinking	Reasoning on the basis of a hypothetical premise	If there were no bees, these crops would not grow
Episodic counterfactual thinking	Reasoning about how specific past events could have turned out differently	If I had not planted these flowers, bees would not have been attracted to my garden
Judgments of type causes[a]	Reasoning about generic categories of causes	Bees pollinate flowers
Judgments of token causes[a]	Reasoning about specific causal events	This bee is pollinating this flower

[a]For a review of the type-token causation distinction relevant to developmental considerations, see Woodward (2011).

reasoning about direct versus indirect outcomes was not replicated in a study by Beck et al. (2010). Despite the inclusion of multiple events in a chain in this previous work, the focus was still on single occurrences involving agents.

Human thought is not focused exclusively on understanding single occurrences involving agents (*token causation*, e.g., "How did I catch the flu?"). We also seek to understand general causal principles in the world (*type causation*, e.g., "How does the flu spread?" and "Why do leaves change colour?"). Understanding how learners approach these types of questions involving type causation arguably involves understanding how they reason about alternatives. For instance, asking "what if the temperature didn't drop?" may inform relevant inferences.

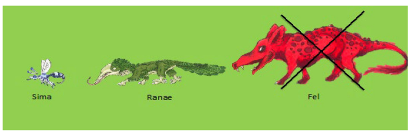
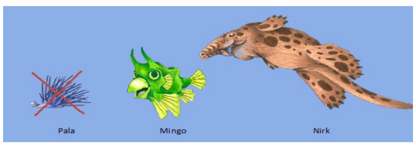
In the current studies, we investigated children's ability to reason about hypothetical changes to a complex and dynamic biological system—food chains. We chose food chains for several reasons. First, they are characterized by interdependence. That is, the survival of each species in a food chain is influenced by the presence of and fluctuations in other species, and thus causal relations are more complex than is typical in developmental studies investigating children's counterfactual reasoning about simple cause–effect relations. In this sense, they are not straightforward, unidirectional causal chains like those in previous studies, and thus existing developmental empirical and theoretical work on counterfactual and causal cognition does not enable clear predictions about reasoning in these cases. Second, food chains involve multiple entities and therefore allow us to look at both direct and indirect effects of hypothetical changes. Third, reasoning about hypothetical changes to food chains and ecosystems may be an important component of understanding more complex concepts such as evolution by natural selection (e.g., if ancestral giraffes had food available at lower levels, would giraffes have long necks today?). Finally, we expected that food chains would be an interesting and motivating topic for children, given the prevalence of animals in children's media, conversations, and classrooms.

Each food chain in the current study was situated in a different biome (e.g., pond, desert) and included three novel animals (e.g., nirks, mingos, and palas). We asked children about the effect of the removal of one of the animals on other animals in the food chain. For instance, in a pond biome, children saw that nirks eat mingos, and mingos eat palas, and were asked to consider what would happen to the other animals if there were no more palas in the pond (Table 2). Children were asked to consider the impact both on proximal and distal animals in the food chain, with the latter requiring several inferential steps. For instance, if there were no palas, there would be no food for the mingos, so there would be fewer mingos. If there were fewer mingos, there would be a shortage of food for nirks, and therefore fewer nirks.

We tested children between the ages of 5 and 7—an age range during which several past studies have found significant developmental changes in episodic counterfactual thinking (e.g., Beck et al., 2006; McCormack et al., 2018; Rafetseder et al., 2010). Children in this age range, especially those in rural and suburban areas, are beginning to understand biological causal relations such as the spread of disease (Coley, 2012). This is also an age when children in many Western countries are first exposed to relevant biological concepts (e.g., animals' diets and habitats) in their classrooms (e.g., Europe: Forsthuber et al., 2011; US: Gillam, 2012; Canada: Ontario Ministry of Education, 2007). Previous research found that with training, 5- to 7-year-olds could generalize concepts (e.g., bigger animals eat smaller animals) from one food chain to another (Vlach & Sandhofer, 2012). Moreover, children of this age are capable of reasoning about complex biological concepts such as evolution by natural selection after training (Kelemen et al., 2014). Evolution as a concept subsumes concepts of fluctuations in food source for species survival. The ability to reason about hypothetical interventions to food chains may therefore be an important piece of reasoning about evolution and other complex biological principles.

If children's reasoning about such phenomena follows a similar timeline to their counterfactual reasoning about token causation involving agents (e.g., Beck

TABLE 2 Test questions and correct answers.

Removed animal	Question type	Example	Correct answer
	Remove top—proximal	"Would there be more, less, or the same amount of ranaes?"	More
	Remove top—distal	"Would there be more, less, or the same amount of simas?"	Less
Bottom prey	Remove bottom—proximal	"Would there be more, less, or the same amount of mingos?"	Less
	Remove bottom—distal	"Would there be more, less, or the same amount of nirks?"	Less

et al., 2006; McCormack et al., 2018; Nyhout et al., 2019; Rafetseder et al., 2010), we may expect children to show significant improvements between the ages of 5 and 7. However, Woodward (2011) has proposed qualitative differences between type and token cause judgments in human reasoning and suggests that children may have *more* difficulty with reasoning counterfactually about single occurrences (token causation; e.g., "How did I get sick?") than generic causal relations (type causation, e.g., "How do people get sick?"), given that children seem to be capable of reasoning with counterfactuals—which underpin the former—later than future hypotheticals—which underpin the latter (e.g., Robinson & Beck, 2000). If type causation is indeed easier, we may expect robust performance across the full age range in the current study. In contrast, we may predict even the oldest children in the present studies will have difficulty reasoning hypothetically about food chains in the absence of scaffolding based on previous findings that: (1) children require training and intervention when reasoning about biological concepts such as food chains and evolution (Kelemen et al., 2014; Vlach & Sandhofer, 2012) and (2) children show a later ability to reason about complex causal models in terms of the number and nature of causal relations present compared to simpler causal models (Nyhout & Ganea, 2021).

Pilot study

We first conducted a pilot study to determine the suitability of our task and age range and to refine our predictions. This study included seventy-two 5-, 6-, and 7-year-old children and is reported in detail in the Supporting Information, but we review the main findings here. The pilot study revealed that the 5 to 7 age range was appropriate, as we saw increasing performance with age, with 5-year-olds performing at chance on all question types, 7-year-olds performing above chance on most question types, and 6-year-olds showing intermediate performance. We found a marginally significant effect of proximity, such that participants had an easier time reasoning about the immediate effects of an animal's removal (i.e., on its direct predator or prey) compared to distal effects (i.e., on its indirect predator or prey). This finding is consistent with German and Nichols' (2003) finding that children were better able to reason about direct than indirect effects of counterfactual changes. We also found a significant effect of question type, such that children had particular difficulty with *distal predator* questions that asked what would happen to the top predator if the bottom prey were removed. We thought this may have been due to the order in which we presented the food chains in the pilot study (always from bottom prey to top predator), and therefore we introduced the variable of *presentation order* to investigate whether this affected performance.

Based on the results of the pilot study, we pre-registered the following predictions for Study 1:

1. Children will show a developmental progression between the ages of 5 and 7, such that performance improves with increasing age.
2. Children will make more accurate inferences when reasoning about proximal (direct effects) than distal animals (indirect effects).
3. In the pilot study, children were better able to reason about the effects of a hypothetical change on distal prey than distal predators. In the current study, food chains were presented in both orders from top predator to bottom prey, as well as bottom to top to ensure there were no order effects. With this change, we predicted that children would show an improved ability to reason about *distal predators* relative to the pilot study.

STUDY 1

Method

Participants

The final sample included 72 children, recruited, and tested at the Ontario Science Centre in Toronto, Canada. Data were collected between July and November 2018. Children were between the ages of 5.01 and 7.97 years: twenty-four 5-year-olds (M=5.48, SD=.29; 12 boys, 12 girls), twenty-four 6-year-olds (M=6.46, SD=.29; 12 boys, 12 girls), and twenty-four 7-year-olds (M=7.44, SD=.30; 12 boys, 12 girls). An additional 4 children were tested but excluded because they did not meet our criteria for English exposure (at least 50% since birth). Demographic information was optionally provided by 61% of families. Parents reported the participating child's ethnicity as White (50%), mixed ethnicity (20%), Chinese (16%), South Asian (5%), Filipino (2%), West Asian (2%), Latin American (2%), and other ethnicity (2%). This sample was representative of the ethnic make-up of the community from which it was drawn. In 80% of families, at least one parent had a bachelor's degree or higher, and in other families the highest level of education attained was reported as community college (14%), high school (2%), and some high school (4%).

Design and procedure

The procedure was pre-registered on the Open Science Framework. Participants sat in front of a computer with an experimenter and were told they would look at different kinds of ecosystems. Children were presented with 4 novel ecosystems, one at a time: a pond, a desert, a forest, and an ocean. Each ecosystem had a linear food chain with 3 novel animals and was presented with illustrated images on a computer (see Table 2). Each animal in a food chain was presented individually with an image on the screen, and the experimenter provided the animal's name and food source (e.g., "This is a mingo. Mingos eat palas."). Children then saw all 3 animals on the screen with their names printed underneath and arrows pointing from prey to predator. This visual display was selected based on consultation with local elementary school teachers and a review of common learning materials.

Children heard the ecosystems presented in one of two orders. In the *bottom-first* order, the experimenter first presented the animal at the bottom of the food chain, then its direct predator, and then the animal at the top of the food chain (i.e., 1-2-3). In the *top-first* order, the experimenter first presented the animal at the top of the food chain, then its direct prey, and then the animal at the bottom of the food chain (i.e., 3-2-1). Children received two ecosystems in the bottom-first and two in the top-first order, and trial order was counterbalanced (either bottom-top-bottom-top or top-bottom-top-bottom).

For each ecosystem, children were first asked a *comprehension* question: "Can you show me who eats whom?" to ensure they remembered the relevant relations between the animals. If children answered the comprehension question incorrectly on their first attempt, the experimenter repeated the predator–prey information until the child was able to correctly demonstrate understanding by verbally responding (e.g., "The nirks eat the mingos and the mingos eat the palas.") or by pointing from each predator to each prey.

Children were then asked two *test questions* for each ecosystem, worded as follows: "Let's imagine if someone had taken all the (palas) away. Would there be more, less, or the same amount of (mingos/nirks)?". This question format was similar to that used by Vlach and Sandhofer (2012) in their training study. Children were asked about the effects of the removal of *one* species per ecosystem on the *remaining two* species. This allowed us to test children's reasoning both about animals above (predators) and below (prey), and directly connected to (*proximal*) and indirectly connected to (*distal*) the removed animal in the food chain. Children received two questions for each ecosystem: one about a proximal animal and one about a distal animal. They were asked about the removal of the bottom prey for two ecosystems (*remove bottom*) and the removal of the top predator (*remove top*) for the remaining two ecosystems. We randomly assigned children to the order of removal of the animals (either *bottom prey, top predator, top predator, bottom prey* or *top predator, bottom prey, bottom prey, top predator*).

Therefore, this study included 3 key within-subjects variables of interest: presentation order (bottom-first or top-first), removed animal (remove bottom or remove top), and proximity (proximal vs. distal). Table 2 displays the dimensions of direction and proximity.

Coding

Response choices (more, less, same) to hypothetical questions were recorded live on paper and later coded. Responses were coded according to accuracy (1 or 0). A second coder watched all 72 recordings and also coded the responses. Table 2 presents correct answers to each question type. Coder agreement was 98% ($\kappa = .95$, $p < .001$). The few discrepancies were corrected by a third coder.

Results

Confirmatory analyses

We conducted confirmatory analyses using a generalized estimating equation (GEE), a semi-parametric regression technique that accounts for covariation between measures in modeling repeated measures or correlated data and is well-suited to binomial data. The model had a binomial probability distribution and a logit link function with age group as a between-subjects predictor, proximity (proximal vs. distal), removed animal (remove bottom vs. remove top), and presentation order (bottom-first vs. top-first) as within-subjects predictors, and hypothetical question score (*0 or 1*) as the dependent measure, because there was one question of each type when taking into account proximity, removed animal, and presentation order.

Consistent with our first prediction, age group was a significant predictor of score, Wald $\chi^2(2) = 14.89$, $p < .001$. We had pre-registered that we would analyze age differences using Mann–Whitney tests but report the more appropriate pairwise comparisons from the GEE here. These comparisons revealed that 6- and 7-year-olds' performance was significantly better than 5-year-olds', $p = .020$ and $< .001$, respectively, but the two older age groups did not differ significantly from one another, $p = .431$. Results of an analysis with exact age as a predictor mirrored these results and are reported in the Supporting Information.

Our prediction that children would make more accurate inferences about proximal compared to distal relations was confirmed. Proximity was a significant predictor, Wald $\chi^2(1) = 31.34$, $p < .001$, such that children's odds of a higher score on proximal items were 14.75 times higher than on distal items, 95% CI [6.62, 32.87].

Removed animal was a significant predictor, Wald $\chi^2(1) = 5.18$, $p = .023$, such that children's odds of a higher score on remove-top items were 9.12 times higher than on remove bottom items, 95% CI [4.10, 20.28]. Recall that we did not make predictions about the effect of removed animal.

Contrary to our third prediction, we did not find an improvement in performance by reversing the order of presentation (bottom-first vs. top-first): presentation order was not a significant predictor, $p = .179$, nor were the interactions between order and the other variables $ps = .449$ to $.452$, suggesting that the difficulty children had with *remove bottom-distal* questions in the pilot study was not a result of the order in which the food chains were presented.

The significance of proximity and removed animal as predictors was subsumed by a significant interaction, indicating that performance differed across the four question types, Wald $\chi^2(1) = 35.07$, $p < .001$. Children's accuracy was significantly lower for *remove bottom-distal* questions than the other three question types, Wilcoxon signed-ranks tests, $ps < .001$. Performance was significantly better on *remove bottom—proximal* than *remove top-proximal* questions, $p = .001$, and on *remove bottom-proximal* than *remove top-distal* questions, $p = .016$. The proportion of children's more, less, or same responses to each question type in each age group are displayed in Figure 1. As is clear in the figure, children's modal response was the correct one for all question types, with the exception of *remove bottom-distal* questions.

Exploratory analyses

To compare children's performance to chance, we conducted exploratory binomial tests comparing the proportion of children scoring 2 out of 2 to chance levels (0.25) for each question type. This is a conservative

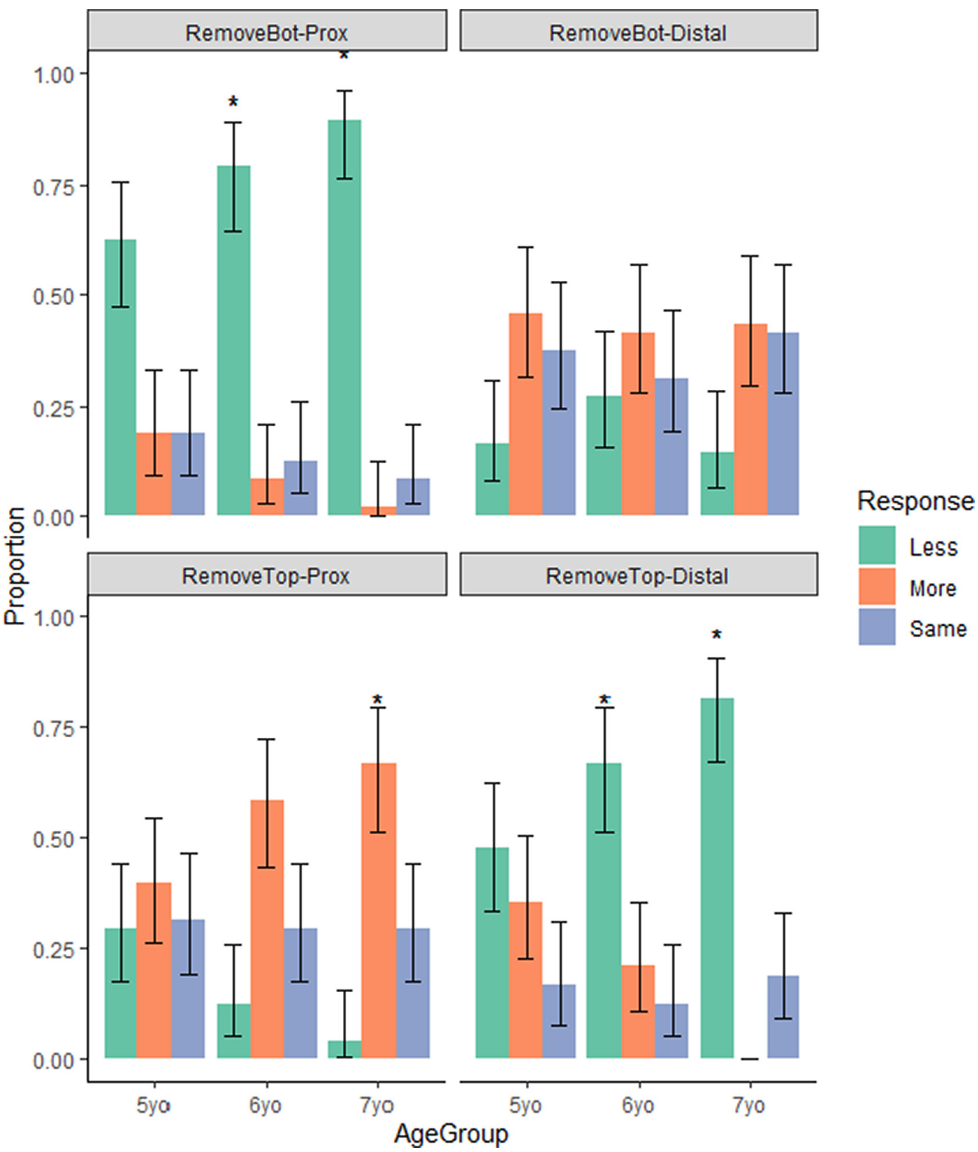


FIGURE 1 Proportion of children's responses that were "less", "more", and "same" in each age group for each of the four question types in Study 1. Children answered 2 questions of each type in this study and therefore proportions are calculated out of a total of 48 responses per question type and age group. Error bars represent 95% confidence intervals. Asterisks denote the proportion of responses that were significantly higher than expected by chance (=0.25), $p < .003$.

chance level, given that there were three response options for each question and children answered two questions of each type. In the following, we corrected for Type I error due to multiple comparisons using an alpha value of .003. For *remove bottom-proximal*, *remove top-proximal*, and *remove top-distal* questions, the proportion of children scoring 2 out of 2 significantly exceeded chance, $p < .001$, whereas the proportion scoring 2 out of 2 for *remove bottom-distal* questions was marginally *lower* than chance, $p = .021$.

We also compared performance against chance for each age group. Five-year-olds' performance was marginally better than chance on *remove bottom-proximal* questions, $p = .007$, but not on the other three question types $ps \geq = .115$, whereas 6- and 7-year-olds' performance was significantly better than chance on all question types, $ps \leq .001$, with the exception of both age groups' performance on *remove bottom-distal* questions, $ps \geq .115$, and 6-year-olds' performance on *remove top-proximal* questions, where performance was only marginally better than chance after correcting for multiple comparisons, $p = .007$.

Discussion

Seven-year-olds, and to some extent 6-year-olds, were able to reason about both direct and indirect effects of a species' removal from a food chain. Although children were not provided with training in the present study, they showed a developing ability to reason hypothetically about these complex and dynamic systems between the ages of 6 and 7.

We found higher accuracy when children were reasoning about direct (proximal) than indirect (distal) effects of hypothetical changes. More specifically, we found that children were more accurate when reasoning about the effects on directly connected predators and directly- or indirectly connected prey than indirectly connected predators. We followed up on and interpret this finding in more detail in the Study 2 Results and Discussion. Presenting food chains from top to bottom or in the reverse order did not appear to affect children's reasoning.

In addition to the differences in reasoning about direct versus indirect effects, we also found significant age differences. Although 5-year-olds were able to correctly indicate which animals eat which on comprehension questions, their performance on hypothetical questions was not robust. They had difficulty with distal questions but showed better success on proximal questions. This finding is consistent with some previous studies showing that children's counterfactual reasoning improves significantly between the ages of 5 and 7 (e.g., McCormack et al., 2018; Nyhout et al., 2019; Rafetseder et al., 2010).

An open question therefore concerns the underlying reason for the developmental differences we observed in this study. The main possibility we consider is that it is the ability to reason through a hypothetical change to a complex and dynamic biological system that develops in this age range. However, there are other possible explanations that do not have to do specifically with the reasoning process.

Five-year-olds may not have sufficient background knowledge and causal understanding about the relation between predators and prey to answer hypothetical questions. Children's biological concepts and ability to reason about biological causation emerges later than their ability to reason in other domains, including physical and psychological causation (Carey, 1985; Hatano & Inagaki, 1994; Wellman & Inagaki, 1997), and reasoning about ecological relations in particular appears to show later development and is related to both to the environment in which children are raised (e.g., urban or rural) and their experience exploring nature (Coley, 2012). Additionally, educational materials on biology shy away from presenting "negative" concepts such as extinction and disease (Shtulman et al., 2021), and as a result children's background knowledge may be lacking in these areas.

Five-year-olds may have struggled to reason within the constraints of the system, perhaps invoking unmentioned factors (e.g., other species, human intervention). This is what researchers have referred to as the nearest possible world constraint: the reasoner should change only those variables that are causally dependent on the hypothetical antecedent, while holding all else constant (Edgington, 2011; Perner & Rafetseder, 2011; Rafetseder et al., 2010). Previous work has suggested that children do not obey this constraint until middle to late childhood in the case of episodic counterfactuals (Rafetseder et al., 2010, 2013), though other work shows reasoning in line with this constraint at age 4 (Nyhout & Ganea, 2019b).

A final possibility is that 5-year-olds may have made different assumptions about the temporal viewpoint from which reasoning should take place. When asked to reason about the implications of the removal of an animal from the food chain, children could have reasoned about the immediate effects on other animals (i.e., little to none) or the longer-term implications (i.e., deaths over time due to unavailability of food source). Our questions asked about the latter, but this may not have been clear to children.

To address these possibilities, we conducted a second study with several changes. We provided more information both to bolster children's background knowledge about the systems and to provide clearer constraints to reduce unwarranted inference. This was done by stating that (1) each species reproduces ("has babies") which increases their population number; (2) a prey species decreases in number when eaten; (3) there are 3 and only 3 animals in each ecosystem, and (4) each animal eats only one other. Comprehension questions were included after each ecosystem to ensure children recognized the key constraints. The background information increased the length of the task significantly, so we only presented children with two ecosystems—one in which they were asked to reason about the removal of the top predator, and one in which they were asked to reason about the bottom prey. We selected two ecosystems (small pond and small island) that were more self-contained to further reduce the likelihood that children would invoke additional factors and to make the fact that each biome housed only three species more believable.

We also revised the wording of the test questions to make the temporal viewpoint clearer by indicating that we were asking about the implications of the hypothetical after "many years." Although children between the ages of 5 and 7 may not understand the precise duration of "many years" (Tillman et al., 2017) we intended to convey that sufficient time had elapsed for the hypothetical changes to play out.

Finally, we asked children to provide explanations for their responses, as we thought these may be illuminating particularly regarding the difficulty children had when reasoning about the effect of the removal of the bottom animal on the top predator in the first study.

We predicted that we would again find age-related increases in performance, and better performance on proximal than distal questions. If 5-year-olds' difficulty in Study 1 was due to a lack of background information, misrepresenting the constraints on the system, or misunderstanding the temporal viewpoint, then we should see improved performance in Study 2 relative to Study 1. If, however, 5-year-olds' difficulty stemmed from difficulties with representing and reasoning hypothetically about a complex and dynamic causal model, then we should *not* see improvements.

STUDY 2

Method

Participants

The final sample included 72 children. Due to the COVID-19 pandemic, children were tested online using video-conferencing software (Zoom) between August 2020 and April 2021. Participants were recruited mainly through our existing database and were residents of the Greater Toronto Area. Additional participants were also recruited through online advertisements and the website Children Helping Science. Children were between the ages of 5.00 and 8.00 years: twenty-four 5-year-olds ($M = 5.43$, $SD = 0.33$; 11 boys, 13 girls), twenty-four 6-year-olds ($M = 6.57$, $SD = 0.34$; 13 boys, 11 girls), and twenty-four 7-year-olds ($M = 7.45$, $SD = 0.34$; 13 boys, 11 girls). An additional 5 children were tested but their data had to be excluded due to issues with the audio on video recordings. Demographic information was completed by 58% of families. Parents reported the children's ethnicity as White (57.5%), mixed ethnicity (19.5%), Chinese (19.5%), South Asian (7%), and Latin American (5%). In most families, at least one parent had a bachelor's degree or higher (85%), and in other families the highest level of education attained was reported as community college (10%) and high school (5%).

Design and procedure

The study procedure was pre-registered on the Open Science Framework. We used only two ecosystems: pond and island, due to the increased length of the task relative to Study 1. Children were asked about the removal of the first animal (bottom prey) for one ecosystem and the third animal (top predator) for the other. Order was fully counterbalanced, and children were randomly assigned to one of four orders based on the combination of ecosystem (pond or island) and removed animal (bottom or top). All test sessions were conducted online, during which the experimenter shared her screen. Because stimuli for Study 1 were presented on a laptop using PowerPoint, the move to the online format for this study did not necessitate a large change in how the task was administered.

A sample script and images for one trial (pond biome—bottom prey removed) are included in the Supporting Information. The procedure included several key changes on the previous study. Additional slides were presented as the experimenter explained the background information about each ecosystem, as indicated in the previous section. This made the task substantially longer, and in piloting we found that four ecosystems were too many for children and they began to lose interest. We therefore reduced the task to only two ecosystems.

After hearing the background on each ecosystem, children were asked the following comprehension questions: (1) "Who eats whom?" (2) "Are there any other animals in the pond/on the island?". During this section, the experimenter also reiterated that each animal could eat only one other, and that there were no other types of animals in the biome.

The test question was also revised from the previous studies to more clearly specify the temporal viewpoint from which the test question was asked: *"Now let's imagine that one day, someone takes away all the palas in the pond. Are you imagining all the palas gone? And let's imagine that we come back to the pond many years after the palas have been gone. When we come back, would there be more, less, or the same amount of mingos?"*

Study 2 included two key within-subjects variables: proximity (proximal vs. distal) and removed animal (bottom prey vs. top predator). Children were asked two hypothetical test questions for each ecosystem (one proximal and one distal) and were also asked to *explain* their answer to each hypothetical question ("*Why would there be more/less/the same amount*?").

Coding

All sessions took place over Zoom and were recorded so children's responses could be later coded offline. Responses were coded on the same criteria as Study 1 according to accuracy (1 or 0). Explanations were coded according to whether they referenced the relevant hypothetical change and the impact it would have on the remaining species (e.g., "Mingos wouldn't have anything to eat." or "The nirks are gone so they aren't eating them anymore."). Correct explanations were given a score of 1 and incorrect explanations were given a score of 0. To gain further insight into children's reasoning strategies, we also scanned explanations to determine the frequency with which children appeared to go beyond the constraints of the system (e.g., by mentioning additional species or causal forces). We also looked at the responses to questions about the effects of removing the bottom animal on the top predator (remove bottom-distal) in more detail to better understand the difficulties children have with these questions.

Results

On the open-ended question asking children who eats whom, responses were highly accurate, with 97.2% and 98.6% of children accurately describing the relations on their first attempt for the pond and island biomes, respectively. When asked whether there were any other animals in the pond or island, 76.4% and 83.3% of children correctly answered "no" for the pond and island biomes, respectively. Children who answered "yes" were reminded by the experimenter that there were only three kinds of animals, and no others. Given the high degree of accuracy, children's responses to comprehension questions are not considered further.

Confirmatory analyses

For the pre-registered main analyses, we used GEEs with a binomial probability distribution and a logit link function. Exact age was a continuous covariate, proximity (proximal vs. distal) and removed animal (bottom prey vs. top predator) were within-subjects predictors, and hypothetical (or explanation) question score (0 or 1) was the dependent measure. We conducted two separate analyses: one for accuracy data and one for explanation data.

Accuracy of responses

Consistent with our predictions and with the results of Study 1, we found that age was a significant predictor of score, Wald $\chi^2(1) = 8.23$, $p = .004$, with improved performance with increasing age.

In contrast to our predictions and Study 1, proximity was not in this case a significant predictor, Wald $\chi^2(1) = 2.24$, $p = .135$. Removed animal was a significant predictor, Wald $\chi^2(1) = 3.90$, $p = .048$, such that children's odds of a higher score on *remove top* items were 4.61 times higher than on *remove bottom* items, 95% CI [2.29, 9.28].

Looking at the four separate question types, the proximity × removed animal interaction was significant, Wald $\chi^2(1) = 18.67$, $p < .001$. Accuracy was significantly lower for *remove bottom-distal* questions than *remove bottom-proximal* and *remove top-distal* questions, $ps < .001$. The difference between *remove bottom-distal* and *remove top-proximal* questions was not significant when corrected for multiple comparisons, $p = .033$, nor was the difference between *remove top-distal* and *remove top-proximal* questions, $p = .034$. The remaining comparisons were not significant, $p = .128$ to .513. The three-way interaction of age group × proximity × removed animal was not significant, $p = .632$.

We compared children's performance to chance for each question type using binomial tests, with chance set to 0.33 because there were three response options (*more*, *less*, or *same*), using a Bonferroni-corrected alpha level of .003. The proportion of children selecting more, less, and same for each question type in each age group is displayed in Figure 2. Children's modal response was correct for all question types, with the exception of *remove bottom-distal* questions and 5-year-olds' responses to *remove top-proximal* questions. Although "same" was never the correct response, children saw this as a valid response option, selecting it on average 33% of the time. We found the same overall pattern of results as in Study 1. Performance was significantly above chance for *remove bottom-proximal*, *remove top-proximal*, and *remove top-distal* questions (all $ps < .001$) but was not significantly above chance for *remove bottom-distal* questions, $p = .142$. We also compared performance to chance for each age group and each question type. Five-year-olds' performance was not significantly better than chance for any question types, $p > .132$. Six-year-olds' performance significantly exceeded chance for *remove bottom-proximal*, $p = .003$ and *remove top-distal* questions, $p < .001$, but not for *remove top-proximal*, $p = .132$ or *remove bottom-distal* questions, $p = .275$. Seven-year-olds' performance was significantly above chance for all questions, $p < .003$, except *remove bottom-distal* ones, $p = .275$.

Explanations

For explanations, exact age was a significant predictor, Wald $\chi^2(2) = 28.04$, $p < .001$, such that children's performance improved with increasing age. Proximity was

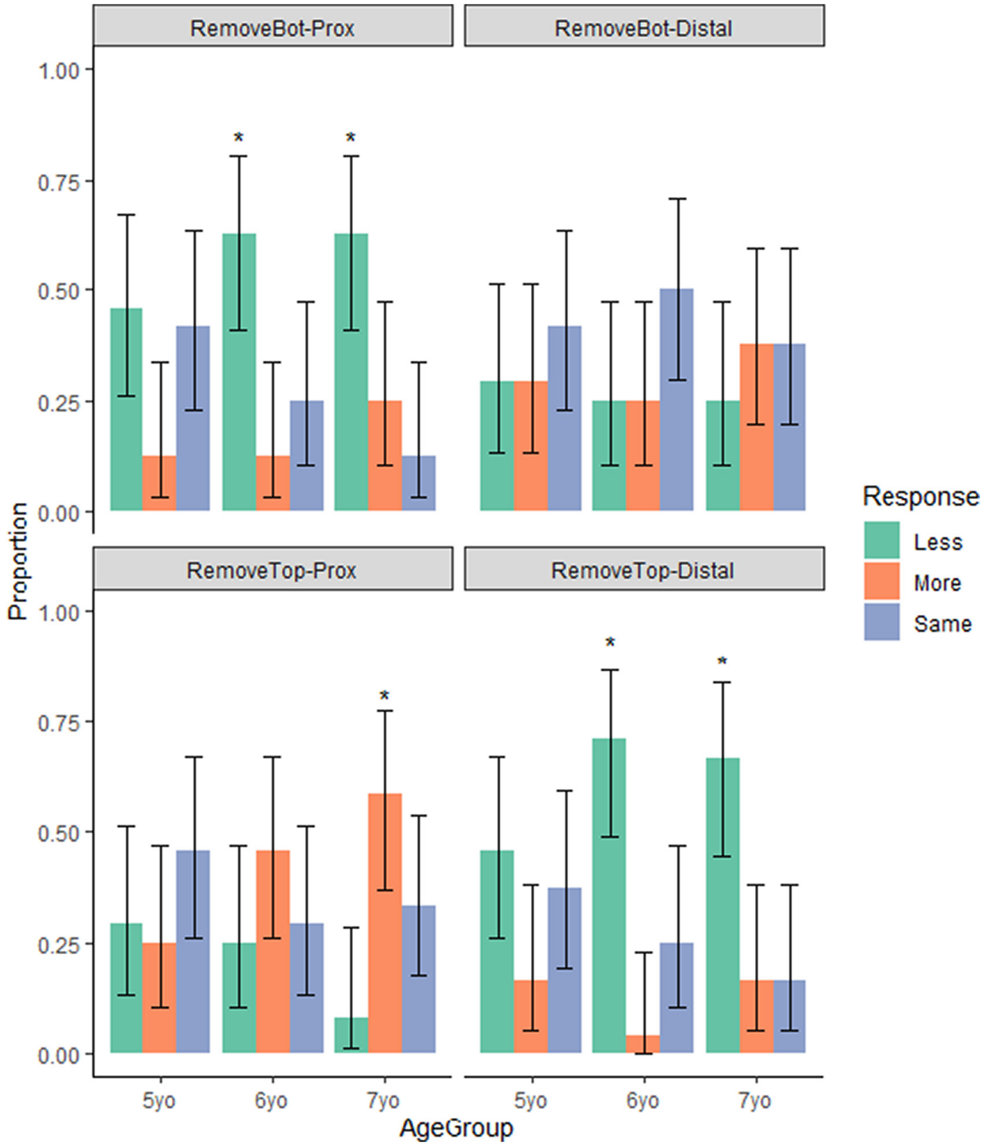


FIGURE 2 Proportion of children's responses that were "more", "less", and "same" in each age group for each of the four question types in Study 2. Children answered 1 question of each type in this study and therefore proportions are calculated out of a total of 24 responses per question type and age group. Error bars represent 95% confidence intervals. Asterisks denote the proportion of responses that were significantly higher than expected by chance (=0.33), $p<.003$.

significant as a predictor, Wald $\chi^2(1)=16.16$, $p<.001$, such that children's odds of a higher score on proximal items (51.4% correct) were 7.66 times higher than on distal items (32.6% correct), 95% CI [1.17, 50.13]. Removed animal was also a significant predictor, Wald $\chi^2(1)=21.65$, $p<.001$, such that children's odds of a higher score on remove top items (54.1% correct) were 11.22 higher times than on remove bottom items (29.9%), 95% CI [4.49, 28.03]. The proximity × removed animal interaction was significant, Wald $\chi^2(1)=15.86$, $p<.001$. Consistent with the previous results of question type, pairwise comparisons revealed that children gave significantly more accurate explanations for *remove bottom-proximal* (47.2% correct), *remove top-proximal* (55.5%), and *remove top-distal* (52.8%) compared to *remove bottom-distal* questions (12.5%), $ps<.001$. All other comparisons were non-significant, $ps=.180$ to .670.

We also conducted *exploratory* analyses of the contents of children's explanations and report descriptive statistics here for some of the main findings. We coded each explanation for whether it referenced predation (being eaten/not eaten), fluctuations in food source, reproduction; invoked additional factors; asserted facts that were already stated; or was uninformative/irrelevant. Cases where children appeared to disobey the constraints of the system (i.e., by invoking additional causal forces) were uncommon. Only 4% (12 out of 288) of children's explanations fell into this category, with most mentioning the existence of humans (e.g., "Because people will catch them with their two hands") which is perhaps not completely unwarranted given that the premise asked them to consider someone taking one species away.

A closer inspection of children's responses to *remove bottom-distal* questions revealed that children may have struggled with these questions because they failed to integrate into their reasoning that the top predator would die of starvation. Of the 72 responses to remove bottom-distal questions, 29% referenced that the top predator would not die because nothing eats it. A further 16% of children answered that the predator would increase in number because they would have babies. Only 11% of children provided correct explanations (e.g., “The fels would have nothing to eat.”). The remaining responses reasserted facts (22.9%; e.g., “nirks eat mingos”), were “I don't know” or irrelevant responses (18.8%) or invoked additional factors (2.7%).

Starvation was the most common explanation offered for *remove bottom-proximal* questions (29% of responses; 50% of 7-year-olds, 33% of 6-year-olds, 4.1% of 5-year-olds). In descending order of frequency, children also provided “do not know” or irrelevant responses (20.8%), mentioned predation (18.1%; e.g., “nirks would keep eating them”), reasserted facts (16.7%), mentioned reproduction (11.1%), or invoked additional factors (4.2%).

Children often explained that the middle animal would increase in number due to decreased predation in response to *remove top-proximal* questions (42.7%), followed by mentions of reproduction (21.1%), reasserted facts (15.9%), “I don't know” (11%), additional factors (8%), and food source fluctuations (1.3%).

Finally, in response to *remove top-distal* questions, children frequently explained that the bottom animal would die due to increased predation (58%). Other responses included “I don't know” (12.5%), reasserting facts (12.5%), invoking additional factors (8.3%), and reproduction (8.3%).

Comparison across studies

In an omnibus GEE comparing performance across studies, we entered study (1 or 2), age group, proximity, and removed animal as predictors, and hypothetical score (proportion correct: 0, 0.5, or 1) as the dependent variable. Study was *not* a significant predictor of score, suggesting that the methodological changes we introduced in Study 2 did not significantly affect performance, Wald $\chi^2(1) = 2.64$, $p = .104$.

We report some of the other major findings here from the omnibus test, given that there were some discrepancies across the two studies. Age group was a significant predictor of score, Wald $\chi^2(2) = 20.28$, $p < .001$, with 7-year-olds and 6-year-olds performing significantly better than 5-year-olds, $p < .001$ and $p = .004$, respectively. Six- and 7-year-olds' performance did not differ significantly, $p = .436$. Proximity was a significant predictor, Wald $\chi^2(1) = 26.29$, $p < .001$, such that children's odds of a higher score on proximal items were 1.75 times higher than on distal items, Exp B = 0.56, 95% CI [0.50, 0.63]. Removed animal was a significant predictor, Wald $\chi^2(1) = 10.07$, $p = .002$, such that children's odds of a higher score on remove top items were 2.23 higher times than on remove bottom items, Exp B = 0.80, 95% CI [0.70, 0.91]. The proximity × removed animal interaction was significant, $p < .001$. Once again, pairwise comparisons revealed that children gave significantly more accurate responses to the remaining three question types compared to *remove bottom-distal* questions, ps < .001. Children also gave more accurate responses to *remove bottom-proximal* than *remove top-proximal* questions, $p = .002$.

Discussion

For the most part, the results of Study 2 echoed those of Study 1. However, unlike Study 1, we did not find a significant effect of proximity in the main analysis. Performance did not differ between Study 1 and 2, indicating that the additional background information we provided did not significantly boost performance. This finding suggests that 5-year-olds, and to some extent 6-year-olds, struggled specifically with the task of reasoning about hypothetical changes to a complex and dynamic biological system, and this is a developing capacity in this age range.

When results were combined across the two studies, we found higher accuracy when children were reasoning about direct (proximal) than indirect (distal) effects of hypothetically removing animals from the food chains. These results are consistent with German and Nichols' (2003) finding that children were more accurate when reasoning about short than long causal chains (but see Beck et al., 2009 for a null result). This effect was mostly explained by the difficulty children had with *remove bottom-distal* questions, which was a robust finding across all age groups and both studies. Children's explanations in Study 2 provide insight into this difficulty. Specifically, a subset of children appeared to fail to make the inference that the top predator would die if its food supply ran out, answering that there would be the same amount of the predator because nothing eats them or there would be more of them because they would have babies. Children seemed to focus on what was *actual* in response to *remove bottom-distal* questions (i.e., lack of predation and reproduction), rather than the hypothetical alternative. We did not specify in the added instruction at the beginning of this study that animals die if their food supply runs out because doing so would have included using a hypothetical in the background knowledge (e.g., “If nirks don't have any food to eat, they will die”) which could have primed the correct answer. We considered this a reasonable trade-off as its inclusion would come with its own problems, even though its lack of inclusion may have led to an underestimation of children's ability on the task. It is worth noting, however, that children seemed to have an understanding of death due to starvation as they readily drew the inference that the middle animal would die if its food supply were gone, and starvation was the most common explanation provided by 6- and 7-year-olds for *remove bottom-proximal* questions. Children's difficulty seemed to be particularly with the distal predator, due perhaps to a combination of the difficulty with drawing indirect inferences and a notion of animals at the top of the food chain as invincible, as media depictions of apex predators (e.g., lions) dying are relatively rare. For instance, in a content analysis of children's books about biology, Shtulman et al. (2021) found that while depictions of predation were common (68% of books), depictions of other negative concepts relevant to species survival such as disease (3%), differential survival (14%), extinction (18%), and competition (10%–12%) were relatively infrequent.

To better understand children's difficulty with remove bottom-distal questions, future studies could present food chains with four species. If children's difficulty stems specifically from reasoning about apex predators, we should expect them to have difficulty reasoning about the animal at the top of the food chain (species 4) but not its direct prey (species 3). If children's difficulty stems from having to make multi-step and indirect inferences, we may also see difficulty reasoning about the animal who is second from the top (species 3). We thank an anonymous reviewer for this suggestion.

GENERAL DISCUSSION

The current project provides novel insight into children's hypothetical reasoning about a complex and dynamic system, and specifically about a biological system. Whereas most previous work has investigated children's episodic counterfactual reasoning in the context of specific past events involving human agents and simple physical causal systems, the current work extends our understanding to *type* causation to study children's reasoning about a complex biological system (i.e., food chains). Across two studies, we taught 5- to 7-year-olds about novel, three-species food chains and asked them to consider the effect of the removal of one of the species on the remaining two. Five-year-olds' performance was at chance across most question types, whereas 7-year-olds were able to reason about both direct and indirect effects of hypothetical changes. Six-year-olds' performance showed some variability across studies, as they reasoned accurately about some question types and not others. Although children were not provided with any training in the present study, they showed an emerging ability to reason about hypothetical changes to dynamic biological systems between the ages of 6 and 7.

What develops?

What develops in this age range, and why did younger children have difficulty making correct inferences in this study? There are a few possible explanations for the developmental progression we observed. The first is that children's ability to reason hypothetically while holding constant the existing causal relations within a system develops in this age range (e.g., respecting who eats whom). A second possibility is that because children's biological understanding is developing in this age range, their hypothetical inferences track with this developing understanding. Younger children may have lacked relevant conceptual knowledge about the causal relations under consideration. Relatedly, younger children may have taken on board the relevant background knowledge about the biological relations but did not truly represent these relations *causally*. A final possibility, not specific to children's domain-specific biological understanding, is that children's ability to adequately represent the causal structure of the event or system under consideration changes in this age range (Nyhout & Ganea, 2019a). The causal structure may have been too complex for younger children to represent, placing too many demands on domain-general executive and processing abilities. We consider these possible explanations in turn.

First, what evidence is there that children failed to respect the constraints or parameters of the system in the current study? Given the frequency of correct answers and the contents of their explanations, older children appeared to respect the constraints by holding constant everything in the causal system and reasoning only about the effects of the hypothetical change (for the related notion of the nearest possible world constraint in developmental research on episodic counterfactual reasoning, see Perner & Rafetseder, 2011; Rafetseder et al., 2010). Even among the youngest children, there were few instances of children mentioning additional animals or auxiliary causes in their explanations, suggesting they attempted to obey the constraints of the systems. Six- and 7-year-olds also answered *distal* questions about the effect of removing the top predator on the bottom prey across both studies at a level well above chance. These questions involved multiple steps of reasoning as the child must first infer and hold in mind what would happen to a proximal animal before making inferences about the distal animal.

Second, do children possess the requisite biological understanding in this age range? We were concerned in Study 1 that children lacked background information from which to draw hypothetical inferences. Supplying additional background information in Study 2 did not lead to an improvement in children's performance, suggesting that it was not a lack of relevant background knowledge that led to difficulty on the task. Younger children appeared to struggle to apply this knowledge to the more complex task of drawing hypothetical inferences. Nevertheless, children's general understanding and knowledge of complex biological causal relations develops through middle childhood, which likely constrains their ability to reason about these relations.

Related to the questions of whether children (1) represented the constraints of the causal system accurately and (2) possessed adequate background knowledge, our task required us to make certain decisions about the information we did versus did not include for children. Even though we taught children what we deemed to be the critical features of the system, we decided not to include more information that could have overwhelmed children's ability to process and retain information. As noted by an anonymous reviewer, certain aspects of the causal system may still have been underspecified (e.g., reproduction rate). While it is possible that children made additional inferences about the food chains that interfered with their ability to make accurate hypothetical inferences, we note that individuals often have to make hypothetical inferences in the face of uncertainty, including in past studies of counterfactual thinking (e.g., Rafetseder et al., 2013).

A third possibility is that even with the necessary background information in hand, 5-year-olds may not have answered hypothetical questions accurately because they did not represent the relevant relations *causally*. Proponents of interventionist theories of causation argue that the ability to reason about hypothetical interventions is what it is to understand a causal system (e.g., Woodward, 2003; see also, the conditional intervention principle, Schulz et al., 2007). On this view, a child who cannot reason about a conditional or hypothetical intervention to a causal relation has not represented the relation in question, as they have failed to represent the counterfactual dependence of the causal relation. Children in our studies answered comprehension questions about which animals eat which others with a high degree of accuracy. There remains the possibility that children may have recognized only descriptively that one animal eats another in the same way that they may recognize, for example, that the family dog eats dog food, without representing the dynamic causal relation between (or mechanism connecting) predator and prey. This is a concern that is present in many previous studies of children's counterfactual and hypothetical reasoning, as it is difficult to measure children's understanding of causal dependence without invoking a hypothetical.

Along with developing biological understanding, a general ability to represent complex causal models develops across this age range with improvements in executive functions and processing capacity. In Study 2, providing more information to constrain the system (e.g., about the fact that each species ate only one other) did not increase performance, and even had the effect of slightly (though not significantly) attenuating it. Given the complexity of the causal models in our ecosystems task, this finding of improvement between the ages of 5 and 7 is consistent with some previous studies indicating that with increasing complexity of causal models (in terms of the number and nature of causal relations between entities), children show later success at counterfactual reasoning across a range of tasks (see Nyhout & Ganea, 2021 for a review). Complex causal models place more demands on children's information processing and executive abilities. Beck and Riggs' (2014) theory of counterfactual reasoning argues that developments in counterfactual reasoning—for instance, advancing from being able to make a counterfactual conditional inference to being able to make inferences about counterfactual emotions—are underpinned by general developments in executive abilities. Indeed, past studies have found significant relations between counterfactual reasoning and executive abilities (e.g., Beck et al., 2009; Beck & Riggs, 2014; Drayton et al., 2011; Guajardo et al., 2009).

Of the candidate explanations we have discussed, the latter two explanations—developments in the nature of children's causal representations and changes in domain-general abilities—are most likely. Both are part and parcel of hypothetical and counterfactual reasoning. Deconstructing these different elements is an important direction for future research.

Future directions

Developmental changes in children's representations of the causal relations in question may drive the developmental differences found in the current study and in other studies of hypothetical and counterfactual reasoning. Despite several theoretical proposals of the integral role of counterfactuals in causal reasoning, many past studies have taken it as a given that children represent causal relationships accurately in counterfactual tasks. If children do not make correct hypothetical or counterfactual inferences, what does this say about their causal representations? Does this necessarily mean that they have not represented relations causally, as asserted by proponents of interventionist theories of causation (e.g., Woodward, 2003)? Understanding the distinct role these representations play in hypothetical reasoning should be a priority in future research.

Given the arguments we have outlined above about the role of domain-general processing and executive function skills in constraining children's ability to reason about complex causal systems, it would also be of interest in future studies to investigate the relation between individual differences in executive functions (e.g., working memory) and performance on tasks requiring children to reason hypothetically about increasingly complex causal models across a variety of domains.

More specifically, it will be important for future research to extend the investigation to a wider variety of tasks and concepts. In the current study, the food chains were isolated, three-species systems to constrain the number of factors children needed to learn about. Future research may investigate how specifying additional details may influence reasoning, as well as how children may reason about more complex real-world systems, both within and beyond the domain of ecosystems.

Finally, future research may also extend this investigation to children from different cultural and geographic backgrounds. The participants in the current study were mainly from middle to upper-middle socioeconomic backgrounds and were residents of a large urban area. This limits the generalizability of the current findings. For instance, in line with previous research finding that children from rural communities engage in more sophisticated biological reasoning than their urban counterparts (Coley, 2012), we may find earlier reasoning about hypothetical interventions to food chains among rural and suburban children.

Significance and implications

The significance of the current findings is two-fold. First, this study indicates that children can reason about hypothetical changes to a complex and dynamic causal system by the age of 7. Reasoning in this age range is not, therefore, limited to counterfactual reasoning about human agents in narratives, nor single occurrences (token causation), as previous work may suggest. Seven-year-olds' ability to answer a variety of questions in the present study indicates that they understand the complex causal structure of these systems. It is an open question—and subject of significant theoretical debate—whether children who failed to answer hypothetical questions correctly necessarily lack a genuine causal understanding of the system in question. More research is needed to study how children's hypothetical reasoning interacts with their causal knowledge in different domains, and how children's reasoning differs across domains.

Second, given children's success at such reasoning by the age of 7, these findings indicate that hypothetical thought experiments may be leveraged in classrooms to teach children complex science concepts—especially in domains in which children cannot feasibly conduct their own physical experiments. However, educators should be mindful of the quantity of information being provided to children, as too much background information could present difficulties. The ability to reason about these problems can be thought of as akin to conducting experiments in the real world (Nyhout & Ganea, 2021) and may lead to meaningful changes in children's understanding in these domains. We asked children about three-species food chains, though future research may investigate whether children are able to reason about more complex problems, such as food chains with competing predators, the introduction of an invasive species, the effects of human intervention or climate change, or the presence of symbiotic relations. These imagined experiments may enable learners to gain a deeper understanding of concepts such as biodiversity, evolution, and climate change. The ability to make predictions about the effects of hypothetical changes to an ecosystem can also provide critical foresight: we can reason about the effects of human-driven ecological changes and biodiversity loss before they unfold.

CONCLUSION

The ability to reason about hypothetical changes to causal systems in the world forms the basis for thought experiments contemplated by scientists in various disciplines. For instance, Gould's "replaying life's tape" thought experiment has inspired theoretical debate and led to controlled "evolutionary re-runs" in laboratories (Blount, 2017). Outside of science laboratories, members of the public contemplate hypotheticals when reasoning about a world without bees or one in which the ice age had not occurred. This type of thought may lead to belief change, behavioral change, and even policy change. The ability to contemplate how things *could be or could have been*—both of the form proposed by Gould and the type engaged by children in the present study—may lead to a better understanding of the world *as it is*.

ACKNOWLEDGMENTS

We would like to thank the research assistants involved in data collection, coding, and stimuli creation, including Lydia Xia, Emily Veall, Mila Milicevic, and Michelle Ou. We would also like to thank the many children and families who made this research possible, and the Ontario Science Centre for facilitating data collection.

FUNDING INFORMATION

This work was supported by a Postdoctoral Fellowship Award from the Social Science and Humanities Research Council of Canada (SSHRC) to A. Nyhout, a SSHRC Insight Development Grant (#430-2019-01128) to A. Nyhout and P.A. Ganea, a Natural Sciences and Engineering Research Council of Canada Discovery Grant (RSD1270194) to P.A. Ganea, and a University of Toronto Excellence Award to H. Sweatman.

DATA AVAILABILITY STATEMENT

The data necessary to reproduce the analyses presented here are publicly accessible. The materials necessary to attempt to replicate the findings are not available publicly due to copyright but are available upon reasonable request to the first author. Analyses were preregistered, except for those indicated as exploratory in the manuscript. Data and the preregistrations for this research are available at the following URL: https://osf.io/nmkts/?view_only=a3678007fb604678a7b8361e9a778e32.

ORCID

Angela Nyhout https://orcid.org/0000-0003-3852-9527
Hilary Sweatman https://orcid.org/0000-0003-1802-2702
Patricia A. Ganea https://orcid.org/0000-0003-0119-7987

REFERENCES

Beck, S. R. (2016). Counterfactuals matter: A reply to Weisberg & Gopnik. *Cognitive Science*, *40*(1), 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., & 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. (2009). Relating developments in children's counterfactual thinking and executive functions. *Thinking & Reasoning*, *15*, 337–354. https://doi.org/10.1080/13546780903135904

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

Blount, Z. D. (2017). Replaying evolution. *American Scientist*, *105*, 156.

Buchsbaum, D., Bridgers, S., Skolnick Weisberg, D., & 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

Carey, S. (1985). *Conceptual change in childhood*. MIT Press.

Coley, J. D. (2012). Where the wild things are: Informal experience and ecological reasoning. *Child Development*, *83*, 992–1006. https://doi.org/10.1111/j.1467-8624.2012.01751.x

De Brigard, F., & Parikh, N. (2019). Episodic counterfactual thinking. *Current Directions in Psychological Science*, *28*(1), 59–66. https://doi.org/10.1177/0963721418806512

Drayton, S., Turley-Ames, K. J., & Guajardo, N. R. (2011). Counterfactual thinking and false belief: The role of executive function. *Journal of Experimental Child Psychology*, *108*(3), 532–548. https://doi.org/10.1016/j.jecp.2010.09.007

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 University Press.

Forsthuber, B., Motiejunaite, A., & de Almeida Coutinho, A. S. (2011). *Science education in Europe: National Policies, Practices and Research*. Education, Audiovisual and Culture Executive Agency, European Commission.

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.

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

Gilbert, J. K., & Reiner, M. (2000). Thought experiments in science education: Potential and current realization. *International Journal of Science Education*, *22*, 265–283. https://doi.org/10.1080/095006900289877

Gillam, D. (2012). A framework for K-12 science education: Practices, crosscutting concepts, and core ideas. *Science Scope*, *36*, 90.

Gould, S. J. (1990). *Wonderful life: The Burgess Shale and the nature of history*. WW Norton & Company.

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*(3), 681–702. https://doi.org/10.1348/026151008X357886

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

Hatano, G., & Inagaki, K. (1994). Young children's naive theory of biology. *Cognition*, *50*, 171–188. https://doi.org/10.1016/0010-0277(94)90027-2

Kelemen, D., Emmons, N. A., Seston Schillaci, R., & Ganea, P. A. (2014). Young children can be taught basic natural selection using a picture-storybook intervention. *Psychological Science*, *25*, 893–902. https://doi.org/10.1177/0956797613516009

McCormack, T., Bramley, N., Frosch, C., Patrick, F., & Lagnado, D. (2016). Children's use of interventions to learn causal structure. *Journal of Experimental Child Psychology*, *141*, 1–22.

McCormack, T., Butterfill, S., Hoerl, C., & Burns, P. (2009). Cue competition effects and young children's causal and counterfactual inferences. *Developmental Psychology*, *45*, 1563–1575.

McCormack, T., Ho, M., Gribben, C., O'Connor, E., & Hoerl, C. (2018). The development of counterfactual reasoning about doubly-determined events. *Cognitive Development*, *45*, 1–9. https://doi.org/10.1016/j.cogdev.2017.10.001

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

Nyhout, A., & Ganea, P. A. (2019b). Mature counterfactual reasoning in 4-and 5-year-olds. *Cognition*, *183*, 57–66. https://doi.org/10.1016/j.cognition.2018.10.027

Nyhout, A., & Ganea, P. A. (2021). Scientific reasoning and counterfactual reasoning in development. *Advances in Child Development and Behaviour*, *61*, 223–253.

Nyhout, A., Henke, L., & Ganea, P. A. (2019). Children's counterfactual reasoning about causally overdetermined events. *Child Development*, *90*, 610–622. https://doi.org/10.1111/cdev.12913

O'Connor, E., McCormack, T., & Feeney, A. (2012). The development of regret. *Journal of Experimental Child Psychology*, *111*, 120–127. https://doi.org/10.1016/j.jecp.2011.07.002

Ontario Ministry of Education. (2007). *The Ontario curriculum grades 1–8: Science and technology*. http://www.edu.gov.on.ca/eng/curriculum/elementary/scientec.html

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. 230–241). Oxford University Press.

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. (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

Robinson, E. J., & Beck, S. (2000). What is difficult about counterfactual reasoning. In P. Mitchell & K. J. Riggs (Eds.), *Children's reasoning and the mind* (pp. 101–119). 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

Shtulman, A., Villalobos, A., & Ziel, D. (2021). Whitewashing nature: Sanitized depictions of biology in Children's books and parent-child conversations. *Child Development*, *92*, 2356–2374. https://doi.org/10.1111/cdev.13571

Tillman, K. A., Marghetis, T., Barner, D., & Srinivasan, M. (2017). Today is tomorrow's yesterday: Children's acquisition of deictic time words. *Cognitive Psychology*, *92*, 87–100. https://doi.org/10.1016/j.cogpsych.2016.10.003

Vlach, H. A., & Sandhofer, C. M. (2012). Distributing learning over time: The spacing effect in children's acquisition and generalization of science concepts. *Child Development*, *83*, 1137–1144. https://doi.org/10.1111/j.1467-8624.2012.01781.x

Weisberg, D. S., & Gopnik, A. (2013). Pretense, counterfactuals, and Bayesian causal models: Why what is not real really matters. *Cognitive Science*, *37*(7), 1368–1381. https://doi.org/10.1111/cogs.12069

Weisberg, D. S., & Gopnik, A. (2016). Which counterfactuals matter? *A response to Beck*. *Cognitive Science*, *40*(1), 257–259. https://doi.org/10.1111/cogs.12241

Wellman, H. M., & Inagaki, K. E. (1997). *The emergence of core domains of thought: Children's reasoning about physical, psychological, and biological phenomena*. Jossey-Bass.

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

Woodward, J. (2011). Psychological studies of causal and counterfactual reasoning. In C. Hoerl, T. McCormack, & S. R. Beck (Eds.), *Understanding counterfactuals, understanding causation: Issues in philosophy and psychology* (pp. 16–53). Oxford University Press.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Nyhout, A., Sweatman, H., & Ganea, P. A. (2023). Children's hypothetical reasoning about complex and dynamic systems. *Child Development*, *00*, 1–16. https://doi.org/10.1111/cdev.13931