Cognition 211 (2021) 104635

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Inhibiting intuition: Scaffolding children's theory construction about species evolution in the face of competing explanations

Samuel Ronfard [a,*], Sarah Brown [b], Erin Doncaster [b], Deborah Kelemen [b]

[a] University of Toronto Mississauga, Mississauga, Canada
[b] Boston University, Boston, USA

ARTICLE INFO

Keywords:
Explanation
Mechanism
Conceptual change
Science learning
Evolution
Coexistence

ABSTRACT

Acquiring the counterintuitive logic of how the mechanism of natural selection (NS) leads to the evolution of new species (speciation) represents a paradigm case of conceptual change. Given this, we examined children's intuitive preconceptions about speciation and their ability to construct, generalize, and retain an accurate understanding of the theory. We did so by conducting two multi-age, multi-session, and multi-measure intervention studies that assessed children's understanding of natural selection 4 times over three months using extended interviews. We also examined the role of Executive Function skills (EF) in these conceptual change processes. Distinctively, we explored whether—consistent with conceptual co-existence accounts—EF not only supports children's initial construction of a counterintuitive theory but also plays an ongoing role in the online reasoning of successful learners. Across two studies, North American children in Grades 2 ($N = 34$) and 3 ($N = 34$) were provided with coherent mechanistic explanations of NS through a two-storybook intervention sequence. The first storybook described the logic of NS to explain how a specialized body part evolved within a fictional species (adaptation). The second storybook extended the logic to explain how this same species evolved into a new, distinct species (speciation). Findings revealed that many second and third graders were able to learn and generalize the logic of speciation. This is a remarkable feat given that speciation conflicts with early developing essentialist and teleological intuitions, and defeats most adults. Our analyses also confirmed that constructing this counterintuitive theory draws heavily on children's EF capacities. They additionally reveal that once the theory is constructed, EF plays a continuing role in reasoning by inhibiting competing intuitive explanations that co-exist rather than being replaced during the process of conceptual change.

1. Introduction

Conceptual change requires children to construct new concepts to express ideas and relations that current conceptual structures do not and cannot represent. There are competing claims about what happens during conceptual change. According to a more traditional constructivist view, concepts and their relationships to each other undergo complete restructuring and prior intuitive ideas are entirely revised and replaced (e.g., Carey, 1985; Gopnik & Wellman, 1994). According to an alternative proposal ("conceptual coexistence") especially when learning deeply counterintuitive concepts, new conceptual structures use old ones as input but are established in parallel with them such that initial conceptual structures are retained. As a result–and consistent with dual-processing theories–they can compete with new ideas during online problem-solving (e.g., Evans, Legare, & Rosengren, 2011; Kelemen & Rosset, 2009; Shtulman & Lombrozo, 2016; for brief overview on dual processing theories, e.g., Evans & Stanovich, 2013).

Prior research with older students and adults suggests that learning about natural selection offers an ideal testing ground for exploring accounts of conceptual change because it is highly counterintuitive: There are important qualitative differences between individuals' intuitive explanations of biological change and a scientifically accurate theory of that same process (Bishop & Anderson, 1990; Shtulman, 2006; Sinatra, Brem, & Evans, 2008). In this paper, we therefore explore conceptual change in relation to the acquisition of the theory of natural selection but we focus on young children rather than adults. Previous work has examined young children's ability to explain natural selection within a species, specifically, in context of trait adaptation–the process of differential survival and reproduction that leads to a species' evolution of specialized body parts (e.g., Kelemen, Emmons, Seston, & Ganea, 2014).

---

* Corresponding author at: University of Toronto, 3359 Mississauga Road, CCT Building, Room 4059, Mississauga, ON L5L 1C6, Canada.
E-mail addresses: samuel.ronfard@utoronto.ca (S. Ronfard), sbrown23@bu.edu (S. Brown), edoncast@bu.edu (E. Doncaster), dkelemen@bu.edu (D. Kelemen).

https://doi.org/10.1016/j.cognition.2021.104635
Received 21 July 2020; Received in revised form 12 February 2021; Accepted 13 February 2021

This research extends this work by examining young children's ability to construct an accurate account of natural selection at the between species level, that is, in relation to the evolution of new species. The process of speciation is likely to be even more difficult for young children to understand than adaptation because speciation conflicts directly with essentialist intuitions that species cannot change (e.g., Evans, 2001; Gelman & Rhodes, 2012). It therefore offers an optimal window for questions about the nature of conceptual change.

In focusing on children's counterintuitive theory construction and its associated processes, we therefore consider a number of issues: First, we document and confirm that children have intuitive preconceptions about speciation before instruction and that these preconceptions are more frequent for speciation than for adaptation (see Brown, Ronfard, & Kelemen, 2020; Emmons, Lees, & Kelemen, 2017). Second, we focus in on children's learning about the evolution of species and explore whether they display capacities to acquire and generalize a basic but accurate mechanistic understanding of speciation based on instruction. Finally, we explore the role of domain-general capacities, specifically executive functions (EF), in supporting conceptual change. In this context, we not only explore the role of EF in children's initial learning and construction of a theory of natural selection but also examine whether EF is implicated in children's expression of scientifically accurate ideas after learning has taken place. That is, we explore whether, among successful learners of natural selection, EF subsequently plays a role in the inhibition of intuitive preconceptions during online scientific reasoning–a pattern that is directly consistent with the claim that intuitive prior conceptions co-exist after theory change rather than being replaced. To preview this research, in what follows, we review work on intuitive (mis)conceptions about natural selection, prior evidence of children's ability to acquire a scientifically accurate understanding of natural selection, and prior indications regarding the domain general capacities implicated in that acquisition.

1.1. *(Mis)conceptions about natural selection*

Natural selection—the process of differential survival and reproduction that occurs over multiple generations—is a foundational scientific concept. It provides an explanation for how members of a species evolve specialized traits (adaptation) with this same mechanism also providing an explanation for how entirely new, distinct species evolve (speciation). However, this population-based process is widely misunderstood by high school students and undergraduates who are the usual recipients of teaching on the topic (Gregory, 2009, for review).

How do individuals misunderstand *adaptation*? Their misconceptions tend to follow a relatively predictable pattern. Adaptation by natural selection is a multi-generational process that occurs due to random phenotypic variation at the level of the population. That is, individual members of the population who happen to have more advantageous traits out-survive and out-reproduce animals with less advantageous traits. As a result, over time, a greater proportion of individuals in the population come to possess these more advantageous traits. Instead of this population- and variation-based account, however, students often conceive of adaptation as a teleological—or need-based—change in an individual's traits (e.g., giraffes acquired long necks because they needed long necks in order to eat). Often, such individual level "transformational" accounts do not elaborate a causal mechanism. Individual organisms simply magically and uniformly change in goal-directed ways to gain traits to serve beneficial purposes (Brown et al., 2020; Shtulman, 2006). Such misconceptions about adaptation by natural selection have their roots in intuitive explanatory biases that are observable from early in development. A key example of one of these is the teleological tendency to view natural phenomena as existing to serve purposes (e.g., Coley & Tanner, 2012; Kelemen, 1999; Kelemen, 2004).

These kinds of early developing explanatory biases seem likely to have an even greater influence on individual children's understanding of *speciation*—how over time selection on multiple traits can lead to the emergence of new species. This is because speciation may be even more counterintuitive than adaptation. In addition to conflicting with teleological intuitions that entities emerge to fulfill purposes, speciation conflicts with our essentialist biases—the tendency to think of species members as possessing an invariant, core property that underlies category membership and the traits associated with it (e.g., Gelman, 2003). This essentialist bias develops early, lessens appreciation of trait variability within a population (Emmons & Kelemen, 2015; Shtulman & Schulz, 2008), and promotes assumptions that category boundaries between species are real and fixed rather than psychologically constructed and biologically flexible (Coley & Tanner, 2015; Gelman & Rhodes, 2012; Kelemen, 2012; Shtulman, 2006). In other words, the cluster of intuitions promoted by inaccurate essentialism effectively shuts down abilities to easily entertain, or accurately represent, the gradual variation-based process by which speciation occurs. Instead, in combination with teleological intuitions, essentialism paves the way, from early in development, for misunderstandings that new species evolve through need-based, species-wide transformational events (Emmons & Kelemen, 2015). Among older students, such misunderstandings are often robust even in the face of instruction. Furthermore, importantly, rather than being revised-and-replaced during the learning process—as suggested by traditional models of conceptual change—studies increasingly suggest that misconceptions about the process of natural selection persist and coexist with formally learned theoretical understandings (e.g., Kelemen, 2012, 2019; Shtulman & Lombrozo, 2016). Thus, one aim of the current study is to document young children's intuitive preconceptions about speciation before and after instruction and compare them to children's preconceptions about adaptation. Given the aforementioned research with adults, we hypothesized that young children would display greater instances of individual, need-based, species-wide transformational events when reasoning about speciation relative to adaptation. Given the co-existence account of conceptual change, we further expected that these intuitions would show signs of persisting alongside accurate understandings after instruction (see Brown et al., 2020; Emmons et al., 2017).

1.2. *Constructing a population-based account of natural selection*

Because intuitive explanatory tendencies (e.g., the teleological and essentialist biases) develop early, it is problematic to delay instruction about natural selection until high school, as is typically the case. In the absence of coherent early mechanistic instruction, intuitively-based misunderstandings have greater opportunities to entrench, and once entrenched, these misunderstandings interfere with initial post-elementary school learning about natural selection (Kelemen, 2019). Questions remain, however, about whether such early coherent intervention is viable when it comes to complex counterintuitive ideas or whether, consistent with conventional wisdom (Metz, 2008), comprehensive transferable learning about a mechanism as challenging as speciation is simply too cognitively difficult for elementary school children.

In pursuit of exploring whether early instruction on natural selection is at all viable, to date, several studies have explored children's abilities to learn about adaptation by natural selection in interventions built around the simple child-friendly pedagogical tool of causal-explanatory picture storybooks (Brown et al., 2020; Emmons, Smith, & Kelemen, 2016, 2017; Kelemen et al., 2014). In these studies, the storybooks—which also form part of the basis for the current intervention—coherently, mechanistically explained adaptation by weaving together the following key concepts: (i) trait variation inherent to a biological population; (ii) effects on habitat and food-sources due to environmental change; (iii) differential health and survival due to differential access to food; (iv) differential reproduction due to differential health; (v) the reliable transmission of stable, heritable physical traits across generations; and (vi) trait-frequency changes (i.e., adaptation) over multiple generations (Kelemen and The Child Cognition Lab, 2017;

Kelemen and the Child Cognition Lab, 2018). For example, the storybook on adaptation called "*How the Piloses Evolved Skinny Noses*", followed the piloses, a fictional anteater species and explained how rising temperatures caused the piloses' insect food to move from above ground into deep, thin underground tunnels. As a result, the rare piloses in the population that had long skinny noses were better able to catch insects than more numerous piloses with wider noses. This led to piloses with thinner noses being healthier, living longer, and reproducing more than piloses with wider noses. Because this process repeated itself over multiple generations, individuals with thinner noses came to predominate although this trait was once infrequent (see Fig. 1a). The gradual nature of adaptation was not only made salient by the text but also in the illustrations that depicted proportional changes in trait frequencies over time on a page-by-page basis.

Intervention studies employing these adaptation storybooks have been effective, with findings revealing that, despite substantive misconceptions at pre-test, 7- to 8-year-old children can acquire a population-based understanding of adaptation by natural selection and generalize it on near and far transfer tests (Brown et al., 2020; Emmons et al., 2016; Kelemen, 2019; Kelemen et al., 2014). This enduring far transfer was particularly apparent after children engaged in a more extended multi-day classroom intervention sequence that involved explicitly analogizing two parallel adaptation storybooks involving selection pressures on food (Emmons et al., 2017). Given this prior success, the present research explores how far children's understanding can extend. Can a multi-day learning sequence help early elementary school students extend the logic of adaptation by natural selection to the theoretically more counterintuitive concept of speciation?

1.3. The role of executive function (EF) skills

In addition to exploring children's learning, we also wanted to examine the cognitive mechanisms that facilitate it. We were particularly interested in the role of EF, which can be divided into three facets: working memory (the ability to hold information in mind), inhibitory control (the ability to inhibit a response), and set shifting (the ability to switch between two rules). The process of conceptual change, whether it is theorized as revise-and-replace or as conceptual coexistence, is thought to involve EF because the construction of new concepts involves the resolution of conflict between existing conceptual structures and the ones being built as individuals reason about the phenomenon under consideration. This process requires all three EF facets because individuals need to hold the competing conceptions in mind (working memory), evaluate each conception while inhibiting the other (inhibition), and toggle back and forth to assess their relative fit (set shifting). Recent research has indeed found that EF predicts children's ability to *acquire* new concepts across a range of STEM-relevant concepts: mathematics (Geary, vanMarle, Chu, Hoard, & Nugent, 2019; Vosniadou et al., 2018), physics (Bascandziev, Powell, Harris, & Carey, 2016), biology (Bascandziev, Tardiff, Zaitchik, & Carey, 2018; see Zaitchik, Solomon, Tardiff, & Bascandziev, 2016 for overview). Thus, one aim of this study was to examine the role of EF in the initial construction of a population-based understanding of natural selection.

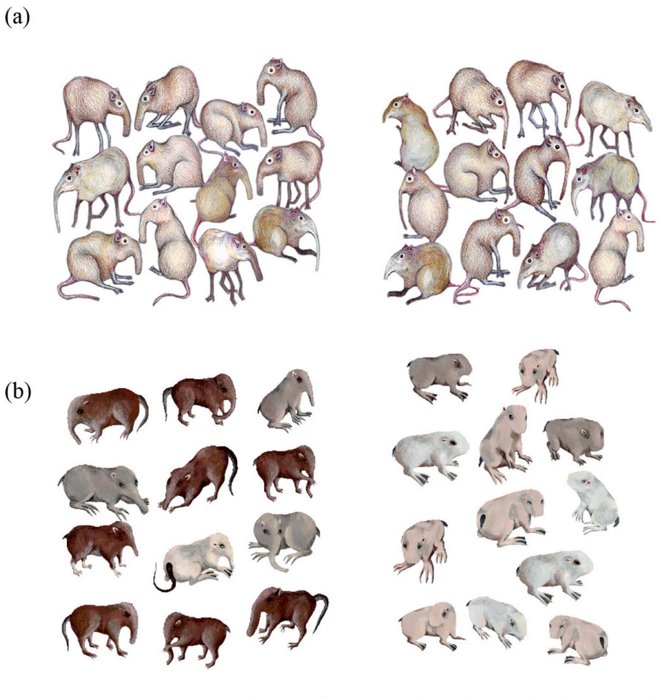


Fig. 1. Representation of (a) the populations of piloses before the change in the climate (left) and multiple generations after the change in climate (right, Book 1) and (b) the initial group of piloses washed away to the island (left) and the group of animals now living on the island after multiple generations who are now called miroungas (right, Book 2).

In addition to this goal, we were also interested in examining the role of EF in inhibiting children's intuitive explanations once they have already acquired a scientifically accurate understanding of natural selection. The prediction that it does is unique to the dual-processing account of conceptual change. According to this theoretical proposal, because intuitive explanations are not replaced but rather persistently coexist alongside formally learned scientific explanations, they need to be inhibited for accurate scientific explanations to be expressed and used during online problem solving. Evidence for this account comes from various studies with adults. These show that: across multiple domains, adults are more likely to endorse intuitive but scientifically incorrect statements under cognitively taxing speeded conditions, even when they are highly scientifically trained (e.g., professors, Kelemen, Rottman, & Seston, 2013; for similar logic see also, for example, Goldberg & Thompson-Schill, 2009; Shtulman & Harrington, 2016); that highly schooled adults who reject intuitive explanations in favor of scientifically accurate judgements have higher levels of inhibitory control (Kelemen & Rosset, 2009); and relatedly that the more frequent intuitive animism of older adults (relative to younger adults) on biological knowledge assessments is a result of declining EF rather than declining or absent biologically accurate conceptual knowledge (Tardiff, Bascandziev, Sandor, Carey, & Zaitchik, 2017). Research with children is more sparse but is generally consistent with conceptual coexistence proposals. Vosniadou et al. (2018) found that children with greater set-shifting and inhibition abilities respond more quickly, and make fewer errors, on tasks pitting intuitive and scientific concepts against one another. However, because children's understanding of the target science concepts was not measured prior to the task, it is not possible to distinguish between children whose greater EF allowed them to construct these concepts during the experiment and children whose greater EF allowed them to express concepts they already possessed.

Thus, in the current study, we focused on the group of children who successfully acquired the theory of natural selection from our intervention. We examined whether their abilities to accurately express that counterintuitive theory at later points (operationalized as not providing an intuitive theoretical misconception) was related to components of EF implicated in inhibitory control. This relationship should be evident only if prior intuitive explanations coexist and compete. To our knowledge, this second analysis is distinctive because, to date, it would provide the most direct test of the claim that EF remains relevant to children's explanations post theory construction.

1.4. The current study

In sum, this study sought to answer four questions to shed light on children's capacities for counterintuitive theory construction about the evolution of species and the cognitive abilities that facilitate that process and subsequent theory mobilization: (1) Prior to instruction on natural selection, do 7–9 year old children hold more preconceptions about speciation than adaptation such that speciation appears more challenging to learn than adaptation?; (2) Does exposure to a storybook intervention sequence designed to scaffold children's coherent accurate understanding of natural selection help children overcome such intuitive explanations and facilitate children's generalizable construction of a population-based account of speciation?; (3) Does EF facilitate children's ability to construct a population-based understanding of natural selection; (4) Once children have acquired such an understanding does EF help children inhibit co-existent and competing intuitive explanations?

To explore these questions, we focused on second graders given their aforementioned ability to learn the process of adaptation from the storybook, "*How the Piloses Evolved Skinny Noses*." To explore their capacity to learn speciation, we designed a new storybook that extended the narrative about the piloses. This new storybook was read after the initial piloses storybook and explained how—after the piloses underwent adaptation of their trunks—geographical isolation and selection on multiple traits led a sub-population of piloses to evolve into the physically dissimilar frog-like "miroungas" species (Fig. 1b). We hypothesized that this sequence of two storybooks would scaffold children's understanding because it would gradually introduce the mechanistic logic of natural selection, first in the easier within-species context of adaptation and then in the more challenging between-species context of speciation.

In Study 1, urban second graders heard the original piloses adaptation storybook and the new miroungas speciation storybook in a four-session sequence that supported children's mapping of within-species adaptation to the larger scale process of speciation (Table 1, each session generally ranged from 40 to 50 mins). Book-readings were interspersed with in-depth talk-aloud assessments that required children to explain adaptation and speciation, applying them to novel scenarios immediately and after 3 months. Assessing children's understanding after a 3-month delay allowed us to document the strength of children's learning and intuitive (mis)conceptions over time.

In Study 2, urban second and third graders completed the same intervention as in Study 1 to assess the reliability of results and to examine developmental trends. In addition, Study 2 added a control group of third grade children to assess the impact of the intervention relative to current instructional practices given recent Next Generation Science Standard-based changes (NGSS) that have led to the introduction of some instruction relevant to evolution by natural selection in third grade (Achieve, 2013; Massachusetts Department of Education, 2016).

Table 1

Participants in Study 1 and Study 2 heard the adaptation storybook and the speciation storybook in a four-session sequence. Book-readings were interspersed with talk-aloud assessments that required children to explain adaptation and speciation, applying them to novel scenarios immediately and after 3 months. EF assessments are described in more detail and analyzed in Study 3.

Session 1 (Pre-test)	**Session 2 (Adaptation)**	**Session 3 (Speciation)**	**Session 4 (Speciation after *3-month-delay*)**
Adaptation Pre-test	Adaptation Storybook Reading	Speciation Storybook Reading	Speciation Delay Novel
Speciation Pre-test	Adaptation Comprehension Post-test	Speciation Comprehension Post-test	Speciation Delay Familiar
	Adaptation Generalization Post-test	Speciation Generalization Post-test	
Acceptance of Common Ancestry Task	Acceptance of Common Ancestry Task	Acceptance of Common Ancestry Task	Acceptance of Common Ancestry Task
			EF Assessments

In addition to the intervention learning measures, children in Study 1 and Study 2 also completed two measures of EF (flanker task and digit span). Thus, following Study 2, we present Study 3 which involves cross-study analyses on the role of EF not only in children's construction of a population-based understanding of natural selection but also in their expression of it once it has been learned.

1.5. Study 1

1.5.1. Data availability

The data and syntax files for Study 1, Study 2, and Study 3 are openly available at the Open Science Framework at https://osf.io/ct8sx/. The adaptation assessments are openly available on the Evolving Minds project website (https://www.evolvingmindsproject.org/materials). The custom storybooks used in the intervention are available online or in bookstores (Kelemen and The Child Cognition Lab, 2017; Kelemen and The Child Cognition Lab, 2020). Additional information about the assessments including the full speciation assessments and how to code them is available in Supplementary Online Materials which are also available at the Open Science Framework link provided above.

1.6. Participants

Our sample size was based on prior research (e.g., Emmons et al., 2017; Kelemen et al., 2014) using similar intervention designs that consistently yielded large effects (pre- to post-test ORs reflecting 12- to 100-fold increases in odds when a large effect is defined as OR > 8-fold increase in odds; Chen, Cohen, & Chen, 2010). Participants were 18 second graders (10 boys, 8 girls, *M* age = 7 years, 10 months, *SD* = 5 months) from three classrooms within a New England urban charter school. Thus, given our repeated measures design, our analyses of adaptation contained 54 data points (18 children × 3 assessments) and our analyses of speciation contained 90 data points (18 children × 5 assessments). One additional second grader was tested, but their data was excluded because of a diagnosed language disability. Age information was missing for two children. Classrooms represented diverse racial, ethnic and socioeconomic backgrounds: 69% of students at the school identified as African American/Black, 24% Hispanic, 3% multi-race or non-Hispanic, 2% Asian, and 2% white. At the time of testing, 72% of students at the school were eligible for free or reduced lunch.

This study was approved by the Ethics Committee of Boston University ("Evolving Minds: Children's learning of biological concepts from picture books", #2350E). Guardians of participants gave informed consent in writing before children participated in the study. Children gave verbal assent.

2. Materials and procedure

2.1. Intervention time table

Participants were tested individually. Table 1 provides a schematic of the intervention and indicates how it built from the theory of adaptation by natural selection to speciation. As Table 1 shows, in *Session 1*, children received two pre-tests, each involving a different realistic but fictional novel species: one assessing their knowledge of adaptation, one assessing their knowledge of speciation. In *Session 2*, children were read the adaptation storybook, "*How the Piloses Evolved Skinny Noses*". Their understanding of natural selection in the context of adaptation was then evaluated in two adaptation post-tests: a comprehension post-test that focused on the piloses and evaluated their understanding of the adaptation book and a near generalization post-test that examined their ability to apply their learning to another novel species that underwent adaptation on a food-relevant trait. Note that the scenarios used for the adaptation pre-test and the adaptation generalization post-test were counterbalanced across children. In *Session 3*, children were read the speciation storybook, "*Meet the Miroungas*". Their understanding of natural selection in the context of speciation was then evaluated twice: a comprehension post-test on the evolution of the miroungas that evaluated children's understanding of the speciation storybook and a generalization post-test that examined their ability to apply their learning to another novel species that underwent speciation due to food-based selection pressures. Note that the scenarios used for the speciation pre-test and the speciation generalization post-test were also counterbalanced across children. *Session 4* took place after a three-month delay. Children completed two speciation assessments. They first completed another speciation generalization post-test with a fictional novel species. In this scenario, a lizard-like species is forced underground as a result of a change in the climate (i.e., shift to scorching temperatures that resulted in advantages for traits facilitating underground burrow living). Children then completed a speciation generalization post-test with a familiar species. This scenario evaluated children's ability to apply the logic of natural selection to explain the evolution of moles from a mammalian ancestor. We reasoned that this last assessment would be hardest for children given that it required them to apply their understanding of natural selection to an animal they were familiar with. In consequence, it would force children to apply their developing understanding of natural selection in a context where they had existing explanatory intuitions as well as existing factual knowledge. At the end of each of the four assessment days, children also completed a four-item task that probed their acceptance of the idea that disparate familiar species can share common ancestry. This common descent task is described in Supplementary Online Materials Appendix 5. On the final day, children also completed EF assessments.

2.2. Storybooks

The two storybooks used in the intervention were custom-made (see Kelemen and The Child Cognition Lab, 2017; Kelemen and The Child Cognition Lab, 2020). We used custom books because, at the time of the investigation, no trade books existed that provided coherent accurate mechanistic descriptions of adaptation and speciation. Designing our own storybooks about novel animals served three purposes: (a) it allowed us to create storybooks that built on one another by continuing the evolutionary story of one species; (b) it allowed us to create simple illustrations designed to support the multi-step causal explanation described in the narrative text and avoid misunderstandings (see Kelemen, 2019); and (c) it allowed us to control for children's baseline knowledge by presenting children with a fictitious mammal species about which they had no prior knowledge. Below we briefly describe the two-storybook sequence (for additional information about the storybooks see Supplementary Online Materials Appendix 1).

Trait Adaptation Storybook. The first book ("*How the Piloses Evolved Skinny Noses*") coherently and comprehensively explained the population-based mechanism of adaptation by weaving together the seven key biological concepts described in the introduction. The anteater-like piloses undergo adaptation of their trunks when climate warming leads their "millibug" insect food to move into deep, thin underground tunnels. Piloses with wider trunks are initially more numerous but, because of their differential foraging advantage in the environment, the rare piloses with longer, thinner noses survive and reproduce more. Over many generations, piloses with thinner noses come to predominate (see Fig. 1a). Note that, in the text, children saw the species name written as "pilosas" rather than "piloses". The spelling changed with trade publication of "*How the Piloses Evolved Skinny Noses*" but research materials retain the original spelling.

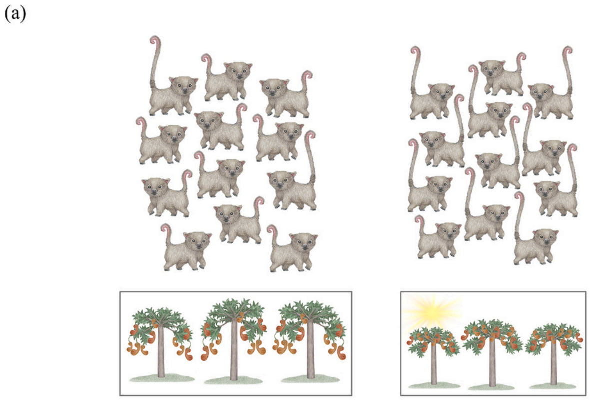
Speciation Storybook. The second book, "*Meet the Miroungas*" continued the story of the piloses. It explained how selection on multiple traits—rather than just one as in the prior adaptation storybook—leads to the evolution of distinct new frog-like mammal ("miroungas"). Specifically, a group of predominantly thinner-nosed piloses is washed away to an island by a deluge of rain. Because the island doesn't have millibugs, the piloses start to forage for an alternative food, calibugs, that can only be found in the water. Because most calibugs swim in the deeper part of the ocean only piloses with shorter tails and noses (that don't drag in the water) and big feet (that can push through more water) are able to catch enough calibugs to be healthy. Piloses without such traits are less healthy and reproduce less. As a result, over multiple generations, the animals on the island evolve into the entirely new frog-like species of miroungas (see Fig. 1b). Indeed, the book explains that, over time, the piloses in the desert and the descendants of the piloses who were washed away to the island become so different that they cannot successfully reproduce with each other and, as a result, are now considered different species with different names.

2.3. Assessments

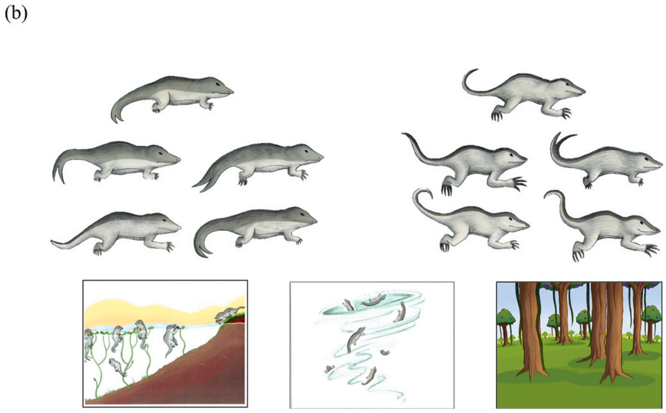
Assessments of adaptation and speciation understanding involved different species but were conceptually similar (see Supplementary Online Materials Appendix 2). The questions used in the adaptation assessments were identical to those in prior work (e.g., Emmons et al., 2016, Emmons et al., 2017; Kelemen et al., 2014, Study 2). The questions used in the new speciation assessments were based on these adaptation questions and followed the same logic. For each type of assessment (adaptation and speciation), children were shown two pairs of images in succession. These two sets of images provided the introduction to the species that was the focus of the assessment and the setup for subsequent questions. The first pair depicted the species in the past (see Fig. 2 for examples). For adaptation assessments, it coupled a picture of the past population in which a particular trait variant predominated (e.g., stumpier tails) with a picture of the species' past habitat and food source (e.g., fruits that grew at the top and bottom of a tree). The second pair depicted the species in the present: its current appearance in which a different trait variant predominated (e.g., stretchier tails) and its post-climate change habitat and food source (e.g., the fruits that now grew only at the top of the tree in a hot, sun-soaked environment).

Speciation assessments had a parallel structure. The first pair of images showed the species in the past in which a particular set of trait variants predominated (e.g., smaller legs, smaller claws, and wider tails) and the species' food source in its past habitat (e.g., vegetation that grew in a pond). The second pair depicted the species' present habitat after a depicted geographical separation event (e.g., a hurricane which displaced a subset of the population to a new habitat). It showed the species' current food source (e.g., vegetation that grew at the top of a tree) and the present appearance of the species in which a different set of trait variants predominated (e.g., bigger legs, bigger claws, and thinner tails). Importantly, in both the adaptation and speciation scenarios, children were never told that the traits of interest had relevance to gaining access to food. They had to infer this relationship when prompted to explain the change in the population.

Following each setup, children answered six closed-ended questions that evaluated their knowledge of isolated facts relevant to understanding natural selection. These questions tapped four concepts: (i) differential survival (2 questions e.g., "Nowadays, will a tardon with a stretchier/stumpier tail probably be healthy and live for a long time?"); (ii) differential reproduction (2 questions, e.g., "Nowadays, will a tardon with a stretchier/stumpier tail probably have lots of children?"); (iii) constancy of traits over the lifespan (1 question, e.g., "See this young tardon. It was born with a stumpier tail. When this tardon is fully-grown, will it be an adult with a stumpier tail or an adult with a stretchier tail?"), (iv) inherited family resemblance (1 question, e.g., "These fully-grown tardons both have stumpier tails. If these two tardons had a child, what kind of tail would their child probably have?"). After giving their initial responses to these closed-ended questions, children were required to justify their answers. Children were only given credit if their closed-ended answer and the justification were correct. Following the six isolated fact questions, children were asked a global open-ended question (e.g., "How do you think that (change in trait frequency) happened?") to probe whether they could self-generate a correct explanation of



Open Ended Question: Many hundreds of years ago most of the fully-grown tardons had stumpier tails but now most of the fully-grown tardons have stretchier tails. How do you think that happened?



Open Ended Question: Many hundreds of years ago, when this group of tegas got carried away to the forest, most of them had smaller legs, smaller claws, and wider tails. Nowadays, the animals that are in the forest–the ones that scientists call iggles– mostly have bigger legs, bigger claws, and thinner tails. How do you think that happened?

Fig. 2. Images used to provide the setup for the (a) adaptation and (b) speciation assessments. For each assessment species and for each type of assessment children were shown the past population (left) and the current (population) as well as the change in habitat (depicted on the second row for each assessment type). Following six closed-ended questions, children were asked the open-ended questions above to assess whether they could self-generate a correct explanation of population change in terms of natural selection.

population change in terms of natural selection. In the case of the adaptation scenario depicted in Fig. 2a, this initial open-ended question was: "Many hundreds of years ago most of the fully-grown tardons had stumpier tails but now most of the fully-grown tardons have stretchier tails. How do you think that happened?". This was followed by prompts for elaboration. These prompts took the form of the experimenter repeating back what the child had already said (e.g., "What happened next after [child's previous response]?"). These prompts were necessary given children's tendency to truncate their answers. Moreover, given that the experimenter never provided new information or corrected children, these prompts made it possible to reveal misconceptions masked by children's unelaborated responses or, alternatively, to uncover their more sophisticated understanding of natural selection. Children also received two further open-ended questions that—with the prompting procedure described above—were designed to elicit further elaboration about the mechanism of change (e.g., "What happened to tardons with stumpier tails?"; "What happened to tardons with stretchier tails?")

For the speciation scenario depicted in Fig. 2b, the basic structure of the six closed-ended isolated fact questions was the same (e.g., "Nowadays, will an iggle with bigger legs, bigger claws, and a thinner tail/smaller legs, smaller claws, and a wider tail probably be healthy and live for a long time?"). The initial global open-ended question was also similar to that used in the adaptation scenario but made explicit mention of the fact that the present population was a different species by using a different name. This made the speciation assessment significantly harder given that the use of the different label for the present population could trigger children's essentialist thinking about species: "Many hundreds of years ago, when this group of tegas got carried away to the forest, most of them had smaller legs, smaller claws, and wider tails. Nowadays, the animals that are in the forest—the ones that scientists call iggles—mostly have bigger legs, bigger claws, and thinner tails. How do you think that happened?" The same prompting for elaboration was used as in the adaptation assessments. Again, children got no feedback on the accuracy of their responses.

2.4. Coding

To facilitate comparisons with prior work, the coding scheme from earlier studies on children's learning of adaptation was used (Emmons et al., 2016; Emmons et al., 2017; Kelemen et al., 2014, Study 2). For each assessment, children were assigned a global score that captured their natural selection understanding based on responses to all isolated fact and open-ended questions. Responses were coded based on a conceptual checklist and a conservative coding rubric. For example, children displaying any evidence of a misconception were never credited with any level of accurate population-based understanding of natural selection (see Supplementary Online Materials Appendix 3, for more information). We coded for four types of misconceptions. Table 2 displays these misconceptions along with definitions and examples.

Children's global understanding of natural selection was classified into one of 5 hierarchical levels. Their understanding was categorized as Level 1, "no isolated facts," when responses to the isolated fact questions demonstrated limited or no knowledge of the prerequisite facts needed to support an understanding of natural selection (i.e., fewer than 5 correct responses to the isolated fact questions). Understanding was categorized as Level 2, "isolated facts but no natural selection understanding," when children displayed robust factual knowledge (i.e., five or more correct responses to the isolated fact questions) but did not generate a correct explanation in response to the open-ended questions or when children displayed a misconception. The three highest levels of

Table 2
Misconception coding scheme with description and examples. Note that developmental and transformation misconceptions are potentially tacitly teleological despite the lack of standard linguistic indicators of teleological explanation (e.g., "so that", "in order to"). Thus, we distinguish them from "explicitly" teleological explanations (see Brown et al., 2020 for discussion of this issue).

Misconception Type	Description	Examples
Developmental	The intuition that an individual member of a species will develop a given, often beneficial, trait as it grows older	(*(The change) happened because) the shorter legs grew up/got older*
Transformation	The intuition that one member/generation of a species is able to spontaneously acquire new (and generally) beneficial traits	(*(The change) happened because) the shorter legs got longer / got stretchier* (no clear mechanism given)
Explicitly Teleological	The intuition that members of a given species develop a trait *in response to a need* for that trait or because traits develop *in order to serve a purpose*	The *wilkies needed longer legs so they got them.* | *The wilkies got longer legs so they could live.* | *The wilkies got longer legs because longer legs helped them get food*
Other	Any other misconceptions	*That's the way they were created* | *The one with the longer legs (lived longer) because he's bigger.*

understanding (Levels 3–5) were only assigned when children demonstrated a robust understanding of the isolated facts (i.e., $\geq$5 correct responses on the isolated fact questions) *and* a correct self-generated explanation of the population-based selectionist logic of natural selection in response to the open-ended questions. A Level 3 or higher categorization was never assigned if there was any sign of a misconception at any point in the assessment. Children were assigned Level 3, "foundation for natural selection understanding," when their open-ended responses included the idea that species members with disadvantageous traits often died while those with advantageous traits tended to survive as a result of selection pressures (i.e., differential survival). They were assigned Level 4, "natural selection understanding in one generation," when they explained the species change in terms of both differential survival and differential reproduction. Finally, children were assigned Level 5, "natural selection understanding in multiple generations," when their open-ended explanations were expanded to explicitly reference the concept that adaptation/speciation occurs over multiple generations.

Two coders coded 100% of the assessments based on transcripts from video recordings. Because pretest and generalization assessments were counterbalanced, coders were unaware of test phase from Session 1 to 3. However, because the delayed test assessments were only used in Session 4, coders may have had some awareness of test phase when coding these particular assessments. Nevertheless, given the number of assessments being coded, and the way coding was organized to keep coders unaware of each individual's learning progress across each test phase of the study (coders did not code each individual on all of their assessments in sequence but rather each assessment using a particular scenario was coded across all individuals), it is likely that this aspect of the design was not particularly salient during coding and thus did not influence coding. Interrater reliability was excellent (Kappa = 0.90), and all disagreements were resolved through discussion.

3. Results

Our analytic process was as follows. When examining whether children possessed more misconceptions about natural selection before instruction in the context of adaptation relative to speciation we used a one-sided test given our theoretically driven directional hypothesis that children would possess more misconceptions about natural selection in the context of speciation than adaptation. When directly examining children's learning, we first examined whether children's performance on our assessments changed over time. This was determined based on the model χ^2. Second, if change was detected, we examined simple effects to understand what changed. To reduce Type 1 error, we adjusted our α from 0.05 to 0.017 in the context of adaptation assessments (Bonferroni's adjustment, 3 post-hoc comparisons: adaptation pre-test vs. adaptation comprehension post-test, adaptation pre-test vs. adaptation generalization post-test, and adaptation comprehension post-test vs. adaptation generalization post-test). In the context of speciation assessments, we adjusted our α from 0.05 to 0.01 (Bonferroni's adjustment, 5 comparisons: speciation pre-test vs. speciation comprehension post-test, speciation pre-test vs. speciation generalization post-test, speciation pre-test vs. speciation delayed novel animal post-test, speciation pre-test vs. speciation delayed familiar animal post-test, and speciation comprehension post-test vs. speciation generalization post-test). We base our statistical conclusions on the aforementioned corrected α levels.

3.1. *Children's (mis)conceptions about natural selection before and after instruction*

Prior to receiving instruction, are children more likely to hold misconceptions about speciation than about adaptation? At pretest, more children offered misconceptions in the context of speciation than adaptation, 89% vs. 67%, and children offered significantly more misconceptions, on average, on the speciation pre-test, than on the adaptation pre-test, $M = 1.22$, $SD = 0.73$ vs. $M = 0.83$, $SD = 0.71$, $t(17) = 1.80$, $p = .045$, one sided test. More specifically, on the pre-test assessments, children were significantly more likely to display transformationist misconceptions in the context of speciation than adaptation, McNemar's χ^2 (1) = 6.40, $p = .011$. There were no statistically significant differences when comparing the prevalence of developmental, explicitly teleological, and other misconceptions across the adaptation and speciation pre-test assessments, McNemar's χ^2 (1) = 1.00, $p = .32$, McNemar's χ^2 (1) = 0, $p = 1.00$, McNemar's χ^2 (1) = 1.00, $p = .32$, respectively.

Misconceptions about adaptation before and after instruction. A majority of children expressed an explanatory misconception during the adaptation pre-test (67%). The percentage of children displaying a misconception fell to 39% during the comprehension post-test and 50% during the generalization post-test (Table 3). However, a repeated measures logistic regression revealed no significant differences in the presence versus absence of a misconception across these assessments, Wald χ^2 (2) = 2.74, $p = .25$.

Misconceptions about speciation before and after instruction. The majority of children expressed a misconception during the speciation pre-test, 89%. Confirming visual inspection of Table 3, a repeated measures logistic regression revealed significant differences in the presence versus absence of a misconception in children's explanations for speciation across the five assessments, Wald χ^2 (4) = 11.83, $p = .019$. Specifically, the percentage of children who displayed a misconception was significantly lower following children's exposure to the speciation storybook on both the speciation comprehension (33%, OR = 0.013) and generalization post-tests relative to the speciation pre-test (50%, OR = 0.04), $p < .001$, $p < .008$, respectively. However, children's ability to respond without stating a misconception faded after a three-month delay. We found no significant differences between the speciation pre-test and the two delayed generalization tests, $p > .01$.

Misconception summary. At pre-test, a large majority of children relied on intuitive but scientifically incorrect explanations to explain natural selection. As expected, they were particularly likely to do so in the context of speciation rather than adaptation scenarios. Moreover, as Table 3 indicates, children found reasoning about speciation to be particularly difficult and, consistent with the challenges of essentialism, they were particularly susceptible to transformationist misconceptions: Children frequently asserted that the new species, and its advantageous form, arose through spontaneous (essential) changes to ancestral species members. Despite these conceptual barriers, exposure to our custom storybooks reduced children's expression of misconceptions about speciation. However, this statistically significant reduction in misconceptions faded after a 3-month delay.

3.2. *Can Grade 2 children acquire an accurate understanding of natural selection?*

Below we report analyses of changes in children's global understanding of natural selection across assessments using repeated measures ordinal logistic regressions. These models examined how the distribution of children across the five hierarchical levels of natural selection understanding changed across time. Odds ratio statistics from these analyses further indicated the magnitude of change in the odds that children's understanding of natural selection improved by one or more levels between two specific assessment times. Preliminary analyses revealed no effect of stimulus order as a result of counterbalancing so we collapsed across them. Fig. 3 displays the change in hierarchical levels of understanding over time.

Natural selection understanding in the context of adaptation. Analyses revealed that children's understanding of adaptation by natural selection changed over time, Wald χ^2 (2) = 14.17, $p < .001$ (See Fig. 3). Relative to the adaptation pre-test, children exhibited a higher level of understanding on the adaptation comprehension post-test following the storybook (OR = 53.05, $p < .001$) and successfully

Table 3

Percentage of second graders (Study 1, $N = 18$) stating each kind of misconception about natural selection at each assessment.

	Any	Developmental	Transformation	Explicitly Teleological	Other
Adaptation					
Pre-test	67%	61%	11%	11%	0%
Comprehension	39%	17%	17%	11%	0%
Post-test	50%	28%	17%	17%	0%
Speciation					
Pre-test	89%	50%	56%	11%	6%
Comprehension	33%	17%	24%	0%	0%
Post-test	50%	17%	28%	17%	6%
Delay novel	56%	28%	33%	6%	11%
Delay familiar	67%	44%	44%	0%	0%

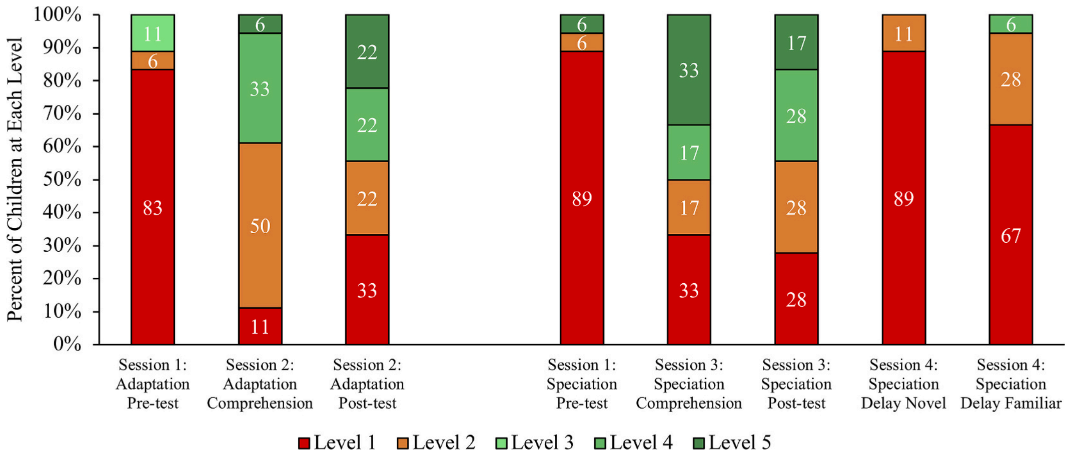


Fig. 3. Results from Study 1: percentages of Grade 2 children who received the Storybook intervention ($N = 18$) classified into the five levels of natural selection understanding for each assessment. Because of rounding, percentages do not always add up to 100. Level 1 = no isolated facts; Level 2 = isolated facts but no natural selection understanding; Level 3 = foundation for natural selection understanding; Level 4 = natural selection understanding in one generation; Level 5 = natural selection understanding for multiple generations.

generalized the logic of adaptation by natural selection to an entirely new animal (OR = 53.16, $p < .001$). Specifically, before hearing the story, only 11% of children displayed a population-based logic (Level 3 or higher). In contrast, on the adaptation generalization post-test, 44% had integrated the facts into an accurate population-based explanation that incorporated, at minimum, the concept of differential survival.

Natural selection understanding in the context of speciation. Analyses revealed that children's understanding of speciation changed over time, Wald χ^2 (4) = 25.77, $p < .001$ (See Fig. 3). Relative to the speciation pre-test, children exhibited a higher level of understanding on the speciation comprehension post-test following the storybook (OR = 68.34, $p < .001$) and successfully generalized the logic of speciation to an entirely new animal on the speciation generalization post-test (OR = 53.93, $p < .001$). Specifically, before hearing the story, only 6% of children displayed a population-based logic on the speciation pre-test (Level 3 or higher). In contrast, by the speciation generalization post-test, 45% had integrated the facts into an accurate population-based explanation of speciation.

However, while many children could learn and apply the logic of speciation, delayed generalization post-tests revealed that children's capacity for generalizing the theory did not endure after 3 months: There was no significant difference in their understanding of speciation when comparing performance on the speciation pre-test and the two delayed generalization tests three months later.

Natural selection understanding summary. Children's understanding of natural selection as a population-based mechanism in the context of adaptation and speciation scenarios increased significantly following their exposure to our multi-day storybook intervention. Despite children's capacities to initially learn and generalize not only adaptation but also speciation, a generalizable understanding of the counterintuitive concept of speciation faded after a 3-month delay.

3.3. Does acquiring a population-based understanding of natural selection in the context of adaptation scaffold children's ability to acquire it in the context of speciation?

Building an effectively sequenced and theoretically coherent learning progression requires establishing that prior learning provides a foundation for later learning. We therefore examined whether children's grasp of the logic of adaptation by natural selection scaffolded their ability to accurately grasp and apply the logic of natural selection when applied to speciation. To do this, we looked at whether children's ability to express a population-based understanding of speciation on the speciation generalization post-test was associated with their ability to generalize adaptation following the adaptation storybook (adaptation generalization post-test): 75% (6 of 8) of children who displayed a population-based logic (Level 3 or higher) on the speciation generalization post-test were previously able to generalize adaptation understanding (i.e., scored level 3 or higher on the adaptation generalization post-test); in contrast, only 20% (2 of 10) of children who did not display a population-based logic on the speciation generalization post-test were previously able to generalize adaptation understanding on the adaptation generalization post-test. Thus, the ability to apply the logic of natural selection to an adaptation scenario appears to scaffold children's ability to apply this logic to a speciation scenario, χ^2 (1) = 5.45, $p = .02$, *Cramer's V* = 0.55.

4. Discussion

We examined children's pre- to post-test learning of natural selection using two measures: reductions in their expression of intuitive but scientifically inaccurate explanations (misconceptions) and their ability to express a scientifically correct population-based account of natural selection. Despite initially high levels of misconceptions about the processes of biological change and little knowledge of the facts or mechanisms of adaptation or speciation, we found that, in the short term, exposure to our intervention reduced children's expression of misconceptions in the context of speciation and increased their understanding of the population-based logic of natural selection in the context of adaptation and speciation. Taken together, these findings demonstrating second graders' capacities to develop a basic understanding of speciation are notable given that this concept is usually not comprehensively or coherently taught until high school. Furthermore, we gained preliminary evidence that a learning sequence on natural selection that built from adaptation to speciation was effective: Children who were able to grasp natural selection in the context of adaptation were more able to grasp it in the context of speciation.

Nevertheless, while our findings revealed that many children could overcome substantive misconceptions to achieve a transferable understanding of speciation, our characterization of it as a deeply counterintuitive concept was confirmed: learning gains decayed over time. After a 3-month-delay, children's expression of misconceptions and ability to generate a population-based account of natural selection to a generalization case no longer differed significantly from pre-test performance levels. Learning and applying the logic of speciation is therefore feasible for early elementary school children but retaining and mobilizing that counterintuitive logic over time is hard.

Given evidence that second graders were able to learn speciation from the storybooks but unable to retain this learning after a three-month delay, in Study 2 we made several changes. First, we modified the 3-month delayed novel animal speciation assessment. During testing and analysis, it had become clear that many children had struggled to infer the functional affordances of the lizard-like species' multiple traits in different habitats, that is, the fitness enhancement conferred by the targeted traits (i.e., shorter spikes, longer claws, flatter bodies) given the environment change (i.e., rising temperatures that led to advantages for traits facilitating underground burrow living). We therefore shifted to a different scenario in which the selection pressure derived from the arrival of predators rather than a change in climate: The advantaged individuals were the rare few whose body color, longer claws and spikes afforded better camouflage and self-defense. Prior research suggests that children have early sensitivities to the latter defensive structure-function relations (Kelemen, Widdowson, Posner, Brown, & Casler, 2003; see supplementary materials for details on assessments).

Second, given evidence that combining assessment and instruction on both adaptation and speciation had been taxing for second graders—overall levels of learning about adaptation were unexpectedly lower than in prior research that has focused on adaptation alone (see Brown et al., 2020; Emmons et al., 2016; Emmons et al., 2017; Kelemen et al., 2014)—we also expanded our age range. In Study 2, we recruited third graders as well as second graders. In context of this change, we also added a control group of third graders who did not receive the intervention. This addition also allowed us to test the effectiveness of the intervention relative to current instructional practices given that, due to alignments with the Next Generation Science Standards (NGSS), instruction that is relevant to understanding evolution by natural selection has recently been introduced at third grade but no earlier (Achieve, 2013; Massachusetts Department of Education, 2016). Adding these two additional groups of third grade children allowed us to address several further interesting questions: First, does the extended storybook intervention lead to different levels (and durations) of change in second versus third grade children's understanding of speciation? Second, do third graders who participate in the storybook intervention show changes in their natural selection understanding that are not seen in third graders who do not?

4.1. Study 2

4.1.1. Participants

Participants were 16 second graders (10 boys, 6 girls, *M* age = 7 years, 7 months, *SD* = 4 months) and 34 third graders (19 boys, 15 girls, *M* age = 8 years, 6 months, *SD* = 4 months) from two second grade and two third grade classrooms within a New England urban public school. These third graders were randomly assigned to the storybook and control condition. Children were drawn from two classes and children within each class were randomly assigned to each condition. One child assigned to the control condition did not complete the two delayed speciation tests. Classrooms represented relatively diverse racial, ethnic, and socioeconomic backgrounds: 66% of students at the school identified as White, 11% Asian, 10% multi-race or non-Hispanic, 9% Hispanic, and 4% African American/Black, and 15% of students at the school were eligible for free or reduced lunch.

This study was approved by the Ethics Committee of Boston University ("Evolving Minds: Children's learning of biological concepts from picture books", #2350E). Guardians of participants gave informed consent in writing before children participated in the study. Children gave verbal assent.

5. Materials and procedure

The materials and procedures were identical to the ones used in Study 1 aside from the change in the delayed novel speciation generalization test described above (see Supplementary Online Materials Appendix 4) and the addition of Grade 3 intervention and control conditions. The same coding scheme was used as in Study 1. As in Study 1, coders were not told the test phase, and to prevent them from tracking individual children's responses over time did not code individuals' assessment in sequence. Coders were also unaware of participants' grade level and condition assignments. Interrater reliability was determined by comparing codes from one coder who completed 100% of the assessments with codes by three other coders (one who completed 50% of the assessments and two who completed 25% each). Interrater reliability between the full coder and the 3 secondary coders was excellent (Kappa = 0.90, 0.91, 0.90). All disagreements were resolved through discussion.

6. Results

We first examined learning for children in the Storybook condition (Grade 2 and Grade 3 combined). We used model χ^2 to determine whether these children's performance on the assessments changed over time. We then examined whether Grade 2 and Grade 3 children in the Storybook condition showed different learning trajectories by testing for the interaction between grade level and assessment. Finally, to determine whether the storybook intervention led to improved performance compared to the business-as-usual control condition, we examined whether Grade 3 children in the Storybook condition experienced greater gains than Grade 3 children in the control condition. When conducting post-hoc tests to examine simple effects we used the same Bonferroni's adjustments as Study 1.

6.1. Children's (mis)conceptions about natural selection before and after instruction

Prior to instruction, are children more likely to hold misconceptions about speciation than about adaptation? Inspection of Table 4 reveals that, as in Study 1, more children expressed misconceptions in the context of speciation than adaptation (82% vs. 70%) and, on average, expressed significantly more misconceptions at pre-test when reasoning about speciation than when reasoning about adaptation, $M = 1.36$, $SD = 0.98$ vs. $M = 1.00$, $SD = 0.90$, $t(49) = 2.31$, $p = .013$, one sided test. Children were also significantly more prone to transformationist misconceptions (ideas that imply notions of essence-

Table 4
Percentage of second graders (Study 2) stating each kind of misconception(s) about natural selection at each assessment.

	Any	Developmental	Transformation	Pure Teleological	Other
2nd Graders Storybook					
Adaptation					
Pre-test	69%	63%	13%	6%	13%
Comprehension	25%	13%	6%	6%	0%
Post-test	25%	19%	0%	6%	0%
Speciation					
Pre-test	88%	56%	44%	38%	13%
Comprehension	19%	13%	6%	0%	0%
Post-test	44%	19%	13%	25%	0%
Delay novel	38%	19%	19%	6%	0%
Delay familiar	63%	13%	38%	13%	6%
3rd Graders Storybook					
Adaptation					
Pre-test	82%	41%	29%	41%	12%
Comprehension	29%	12%	24%	12%	0%
Post-test	18%	6%	6%	6%	0%
Speciation					
Pre-test	82%	35%	65%	35%	24%
Comprehension	24%	0%	12%	12%	0%
Post-test	24%	12%	18%	12%	6%
Delay novel	35%	6%	0%	24%	6%
Delay familiar	35%	0%	24%	18%	6%
3rd Graders Control					
Adaptation					
Pre-test	59%	12%	24%	29%	18%
Comprehension	71%	12%	41%	29%	12%
Post-test	47%	18%	12%	24%	6%
Speciation					
Pre-test	76%	6%	41%	41%	12%
Comprehension	59%	0%	41%	24%	6%
Post-test	41%	6%	29%	18%	18%
Delay novel	25%	13%	6%	12%	0%
Delay familiar	56%	6%	44%	12%	0%

changing) when explaining speciation rather than adaptation, 50% vs. 20%, McNemar's χ^2 (1) = 8.91, p = .003. There were no statistically significant differences when comparing the prevalence of developmental, explicitly teleological, and other misconceptions across the adaptation and speciation pre-test assessments, McNemar's χ^2 (1) = 1, p = .32, McNemar's χ^2 (1) = 2, p = .16, McNemar's χ^2 (1) = 0.14, p = .71, respectively.

Misconceptions about adaptation before and after instruction. Overall, there was a main effect of assessment for the children in the Storybook condition (Grade 2 and Grade 3 combined), Wald χ^2 (2) = 15.18, $p < .001$. Children showed a marked reduction in misconceptions after the intervention: While 76% of children in the Storybook condition displayed misconceptions on the adaptation pre-test, only 21% did so on the adaptation generalization post-test, OR = 0.02, p < .001.

There were no developmental differences in the effectiveness of the storybook intervention as indicated by the absence of a significant interaction between grade level and assessment (pre-test, comprehension post-test, generalization post-test), χ^2 (2) = 1.33, p = .51. Indeed, children in both grade levels showed marked reductions in their misconceptions across assessments (Grade 2: Wald χ^2 (2) = 6.52, p = .038, Grade 3: Wald χ^2 (2) = 8.13, p = .017).

In contrast to children in the intervention conditions, Grade 3 Control children did not experience a reduction in their misconceptions, the level of which remained unchanged over the three adaptation assessments, Wald χ^2 (2) = 2.45, p = .29. Furthermore, the interaction between receiving the storybook intervention and the three assessments was significant, χ^2 (2) = 8.82, p = .012, confirming that the Storybook intervention led third grade children to display different patterns of change in their expression of misconceptions compared to Grade 3 children in the control condition.

Misconceptions about speciation before and after instruction. There was a main effect of assessment for the children in the Storybook condition (Grade 2 and Grade 3 combined), Wald χ^2 (4) = 25.38, p < .001. Relative to the pre-test, children in the Storybook condition experienced a significant drop in misconceptions on the comprehension, 85% vs. 21%, OR = 0.014, $p < .001$, and generalization tests, 85% vs. 33%, OR = 0.034, $p < .001$, immediately after the speciation storybook. This significant reduction in misconceptions persisted after a three-month delay on both generalization post-tests, OR = 0.041, $p < .001$, OR = 0.084, p = .001, respectively. For example, after the three-month delay, the percentage of Storybook children with misconceptions on the final post-test (48%) was significantly lower relative to the pre-test (85%).

Mirroring the adaptation results, the Storybook intervention was similarly effective at reducing misconceptions about speciation for Grade 2 and Grade 3 children—a grade level by assessment interaction was non-significant, χ^2 (4) = 2.97, p = .56. Indeed, children in both grade levels showed marked changes in their expression of misconceptions over time (Grade 2: Wald χ^2 (4) = 13.38, p = .01, Grade 3: Wald χ^2 (2) = 13.07, p = .011).

In contrast to children in the intervention conditions, Grade 3 Control children did not experience a reduction in their misconceptions. The level of these remained unchanged over the five speciation assessments, Wald χ^2 (4) = 9.43, p = .051. However, although Grade 3 who received the storybook intervention displayed a decrease in misconceptions over the course of the study while their counterparts in the control condition did not, the interaction between receiving the storybook intervention and the five assessments was not significant, χ^2 (4) = 6.22, p = .18.

Misconception summary. At pre-test, a large majority of children relied on intuitive but scientifically incorrect explanations to explain natural selection. As in Study 1, participants were more likely to have misconceptions about speciation than adaptation. Nevertheless, exposure to our custom storybooks reduced children's expression of misconceptions and this effect was robust. It was observed following children's exposure to the storybooks and after a three-month delay. No changes in the presence of misconceptions were seen in Grade 3 Control children who were not exposed to the storybook intervention. Tests of the condition by assessment interaction for Grade 3 children confirmed significant differences in children's expression of misconceptions on the adaptation but not on the speciation assessments further confirming the highly counterintuitive nature of speciation.

6.2. Can Grade 2 and 3 children acquire an accurate understanding of natural selection?

Natural selection understanding in the context of adaptation.

Analyses revealed that the understanding of natural selection changed significantly for children in the Storybook condition (Grades 2 and 3 combined), Wald χ^2 (2) = 30.25, $p < .001$, See Fig. 4. Relative to the pre-test, children in the Storybook condition exhibited a higher level of understanding on the comprehension post-test after the storybook, OR = 39.95, $p < .001$, and successfully generalized to a completely novel animal as well, OR = 23.20, $p < .001$. Specifically, before hearing the story, only 9% of children displayed a population-based logic (Level 3 or higher) but 64% did so on the adaptation generalization post-test.

Importantly, the Storybook intervention was similarly effective at

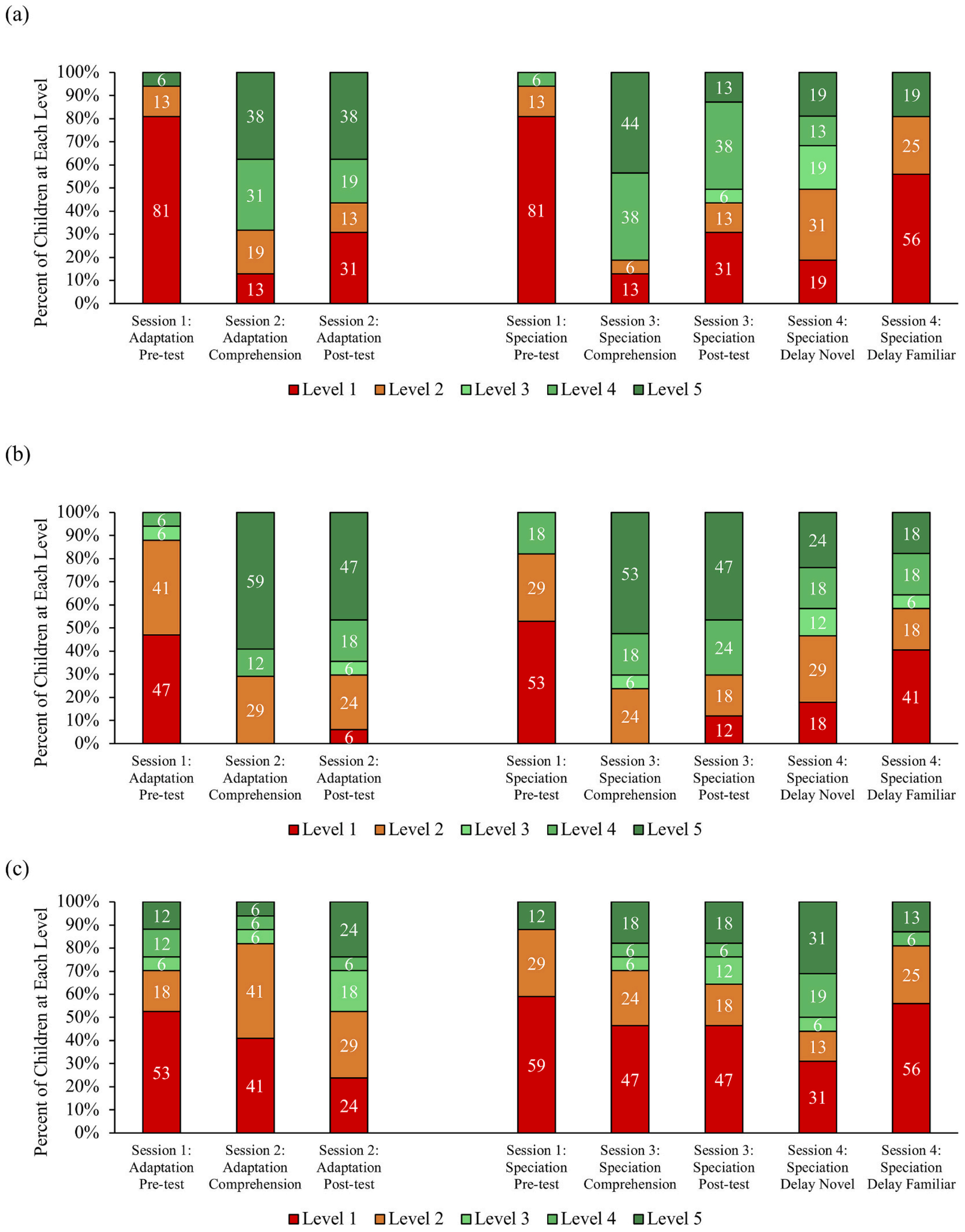


Fig. 4. Results from Study 2: (a) percentages of Grade 2 children who received the Storybook intervention (N = 16); (b) percentages of Grade 3 children who received the Storybook intervention (N = 17); (c) percentages of Grade 3 children in the control condition (N = 17) classified into the five levels of natural selection understanding for each assessment. Because of rounding, percentages do not always add up to 100. Level 1 = no isolated facts; Level 2 = isolated facts but no natural selection understanding; Level 3 = foundation for natural selection understanding; Level 4 = natural selection understanding in one generation; Level 5 = natural selection understanding for multiple generations.

improving Grade 2 and 3 children's understanding as indicated by the lack of a significant grade level by assessment interaction, χ^2 (2) = 0.62, $p = .73$. Indeed, children in both grade levels showed marked improvements in their understanding of adaptation (Grade 2: Wald χ^2 (2) = 12.06, $p < .01$, Grade 3: Wald χ^2 (2) = 17.61, $p < .001$).

In contrast to children in the intervention conditions, despite a significant model test, Wald χ^2 (2) = 6.30, $p = .043$, post-hoc tests revealed no significant changes in Grade 3 control children's understanding of adaptation. Furthermore, the interaction between receiving the storybook intervention and the three assessments was significant, χ^2 (2) = 12.34, $p = .002$, confirming that the Storybook intervention led Grade 3 children to display different patterns of change in their understanding over time compared to the children in the control condition.

Natural selection understanding in the context of speciation. Analyses revealed that children in the Storybook condition (Grade 2 and 3 combined) experienced significant change in their understanding of speciation across the five assessments, Wald χ^2 (4) = 49.16, $p < .001$ (See Fig. 4). Relative to the pre-test, children in the Storybook condition improved their understanding of natural selection on the speciation comprehension post-test, OR = 56.33, $p < .001$, and on the immediate speciation generalization post-test, OR = 20.41, $p < .001$. At pre-test 12% of children displayed a population-based logic (Level 3 or higher) while 64% did so on the speciation generalization post-test. Children's learning partially endured. It was still evident three-months later on the novel animal scenario, 52% displayed a population-based logic versus 12% at pre-test, OR = 11.63, $p < .001$. However, children's performance on the familiar animal scenario did not differ significantly from their pre-test performance, 30% displayed a population-based logic versus 12% at pre-test, OR = 3.58, $p > .01$.

A non-significant grade level by assessment interaction, χ^2 (4) = 3.48, $p = .48$, indicated that the Storybook intervention was similarly effective at improving Grade 2 and Grade 3 children's understanding. Indeed, children in both grade levels showed marked changes in their understanding of speciation over time (Grade 2: Wald χ^2 (4) = 25.57, $p < .001$, Grade 3: Wald χ^2 (4) = 24.75, $p < .001$).

In contrast with Grade 3 children in the intervention condition, Grade 3 children in the control condition did not exhibit a clear pattern of performance or any consistent learning and generalization. Although the omnibus test indicated an effect of assessment for this group, Wald χ^2 (4) = 9.57, $p = .048$, post-hoc tests revealed no statistically significant differences in their understanding of speciation between the pre-test and the comprehension or generalization post-test, or the pre-test and delayed familiar animal post-test. However, control children did display a higher level of understanding of natural selection on the delayed novel animal post-test relative to their pre-test, OR = 11.13, $p = .004$. This surprising significant difference on one test may reflect specific aspects of this assessment given that the structure-function relationships had been changed from Study 1 so that they were especially transparent. As we consider later, some specifics of NGSS-based science instruction intended for third graders may also have played a role.

Furthermore, the interaction between receiving the storybook intervention and the five assessments was significant, χ^2 (4) = 10.96, $p = .027$, confirming that the Storybook intervention led children to display different patterns of change in their understanding over time. Specifically, relative to Grade 3 control children, Grade 3 children who received the storybook intervention experienced significantly greater amounts of change between the speciation pre-test and the speciation comprehension post-test, $p = .003$, and significantly greater change between the speciation pre-test and the speciation generalization post-test, $p = .009$. However, Grade 3 storybook children's speciation generalization performance dropped on the generalization post-tests after a 3-month delay. As a result, we found no difference between these two groups of Grade 3 children when comparing their performance on the two delayed post-tests.

In sum, children in the storybook condition experienced greater pre- to post-test change in their understanding of speciation immediately following the receipt of the storybook intervention relative to children in the control condition. However, after a three-month delay, there was no statistically significant difference between children's speciation pre-test and the two delayed speciation generalization assessments for Grade 3 children in the Storybook and Control conditions.

Natural selection understanding summary. Consistent with the results of Study 1, Grade 2 and Grade 3 children's understanding of natural selection as a population-based mechanism in the context of adaptation and speciation scenarios increased significantly following their exposure to the storybook intervention. Importantly, Grade 3 children who did not receive the storybook intervention did not experience consistent patterns of changes in their understanding of natural selection. In fact, comparisons between Grade 3 children who did and did not receive the storybook intervention revealed that children exposed to the storybook intervention had significantly greater changes in understanding natural selection on both adaptation and speciation post-tests that tested their ability to express their knowledge immediately following the interventions. However, after a three-month delay, this difference between conditions dissipated.

6.3. Does acquiring a population-based understanding of natural selection in the context of adaptation scaffold children's ability to acquire it in the context of speciation?

To examine whether understanding adaptation by natural selection scaffolded children's ability to understand speciation, we looked at associations between the adaptation generalization post-test and the speciation generalization post-test: 78% (21 of 27) of children who displayed a population-based logic (Level 2 or higher) on the speciation generalization post-test had previously displayed this understanding on the adaptation generalization post-test; in contrast, only 30% (7 of 23) of children who did not display a population-based logic on the speciation generalization post-test had previously displayed this understanding on the adaptation generalization post-test. Thus, replicating Study 1, we find evidence for the efficacy of this learning sequence: Learning within-species adaptation by natural selection appears to scaffold children's understanding of speciation, χ^2 (1) = 9.43, $p = .002$, *Cramer's V* = 0.43.

7. Discussion

Findings from Study 2 provide further support for the proposal that, contrary to conventional wisdom, young children are able to learn the complex concepts of both adaptation and speciation by natural selection. As in Study 1, we found that learning the logic of natural selection in the context of adaptation scenarios scaffolded children's ability to apply this same logic to speciation scenarios. With regards to understanding adaptation and speciation by natural selection, we found that Grade 2 children are able to acquire these concepts as well as Grade 3 children. Indeed, on the re-designed delayed novel animal post-test of speciation, many still displayed a significantly higher understanding after a three-month delay, perhaps because the assessment had particularly clear structure-function relationships. By contrast, consistent gains were not seen in the business-as-usual control sample of Grade 3 children despite an anomalously strong performance on one test—the re-designed delayed novel animal post-test involving a predation scenario. One explanation for this is that the delayed post-test coincided with the Grade 3 life science unit. Because testing took place over several days, it is not possible to know what science lessons each student had received, but it is possible that some or all children had started to be exposed to instruction relevant to—albeit not on—the mechanism of natural selection (e.g., how certain traits increase an individual's chance of survival, for example, from predation;Massachusetts Department of Education, 2016). Even as this distinctive stronger performance might have been partly attributable to recent specific changes in Grade 3 business-as-usual science curriculum standards, significantly greater improvements on the speciation comprehension and generalization post-test assessments for Grade 3 children who received the storybook intervention relative to those who did not demonstrates that the storybook intervention generates greater gains than current instructional approaches. However, differences between children in the Grade 3 Storybook and Control conditions while still visible were no longer statistically significant after a 3-month delay. Again, this underscores the challenge of constructing an understanding of natural selection in the context of speciation and suggests that children who received the storybook intervention would probably benefit from additional instruction to solidify their understanding, for example, another storybook on speciation.

7.1. Study 3

We now turn to focus on children's executive function performance and the mechanisms associated with children's construction and expression of the theory of natural selection. To do this, we pooled together data from Studies 1 and 2 to address two questions: First, what role did EF capacities play in children's abilities to overcome intuitive preconceptions and construct an initial accurate understanding of natural selection based on exposure to the adaptation storybook? Second, among children who successfully constructed such a population-based understanding, were greater inhibitory EF capacities also implicated in their explanations about natural selection after a three-month delay given the potential need to manage interfering competition from persistent coexistent misconceptions. To answer the first question, we focused on the role of EF in relation to children's performance on the first assessment of adaptation understanding following the intervention: the comprehension post-test. To answer the second question, we focused on the role of EF in relation to children's performance on the final delayed speciation post-test (the familiar animal post-test)—an assessment that was conducted after a 3-month delay and therefore represented the strongest test of children's abilities to maintain and marshal their learning.

7.2. Participants

A sample of 46 children was created by combining the data from second and third graders who received the adaptation storybook intervention in Studies 1 and 2. Due to missing birth date information on two children in Study 1 (repeated attempts to obtain this information were unsuccessful), this sample was reduced to 44 children (22 girls) when controlling for age, *M* age = 8 years, *SD* = 7 months.

8. Measures

8.1. Measures of executive functions

Children completed two executive function tasks at the end of the final assessment day: a measure that assesses all three facets of EF: working memory, inhibitory control, and cognitive flexibility (the flanker task) and a working memory task (forward and backward digit span). In the Flanker task (Davidson, Amso, Anderson, & Diamond, 2006; Diamond, Barnett, Thomas, & Munro, 2007), the stimuli are lines of fish. Children complete a block of trials in line with an initial rule about the fish ("when the fish are blue, feed the middle fish by pressing the button that corresponds to where the middle fish is facing i.e., the left button if facing left; the right button if facing right). Then, children complete a block of trials following the second rule (when the fish are pink, the task is to feed the fish that are on the outside, feed the fish on the outside by pressing the button that corresponds to where the outside fish are facing). Finally, in the third and final block of randomly mixed trials, children are instructed that they will play both rules at the same time. Consistent with Bascandziev et al. (2018), we used accuracy on the mixed trials as the measure of set shifting and inhibitory control capacities.

To measure working memory, children completed the two parts of the digit span task. The first part probed forward digit span: Children were asked to listen to a set of recorded digits (e.g., 3–8–6) and then repeat back the sequence. They began with a three-digit sequence. The longest possible sequence consisted of 9 digits. The second part was the backwards digit task in which children were told to listen to a sequence of numbers and then recount the sequence in backwards order. They began with a two-digit sequence. The longest possible sequence consisted of 8 digits. For the forward and backwards digit span task, children completed two trials for each sequence. Children's score was the highest sequence they completed. Children failed a sequence when they made an error in recalling digits on the two trials making up each sequence length. The sum of correct trials on the forward and backwards task was our working memory measure.

8.2. Measures of natural selection understanding

Participants' performance on the assessments from Study 1 and 2 were used. We generated a new variable that reflected children's pre- to post-test construction of a selectionist understanding: Children scored 1 if they did not understand adaptation by natural selection at pre-test (i.e., their global understanding score was below 3) but did have it at the adaptation comprehension post-test, ($n = 28$). In contrast, they scored 0 if they did not have the understanding at either the pre-test or comprehension post-test ($n = 18$).

9. Results

9.1. Do executive functions facilitate children's acquisition of an accurate understanding of adaptation by natural selection following exposure to the storybook intervention?

To assess whether—in context of frequent intuitive preconceptions—children's EF skills facilitated their ability to construct an understanding of adaptation by natural selection after exposure to the storybook, we examined whether children's ability to construct the theory between the pre-test and the comprehension post-test differed based on their EF scores (controlling for age).

We tested a series of logistic regression models using the two measures of EF as predictor variables. Given co-existence (dual processing-based) accounts of conceptual change, we expected inhibition to be a particularly strong predictor of children's ability to suppress intuitive pre-conceptions and thus express and learn about natural selection as a population-based mechanism. As a result, we expected that total accuracy on the mixed trial of the flanker task would remain a significant predictor when controlling for working memory and age, thereby providing support for the claim that set shifting/inhibition, in particular, is involved in conceptual change. In Table 5, we display the series of models we used to test this hypothesis. We found that when controlling for working memory and age, children's total accuracy score on mixed trials of the flanker task was a significant predictor of their ability to construct a population-based understanding of natural selection following their exposure to the adaptation storybook, OR (for a 10-percentage point increase in EF accuracy) = 1.76, $z = 2.15$, $p = .032$.

9.2. Do executive functions remain relevant to children's ongoing expression of their understanding of natural selection in the context of speciation by inhibiting co-existent misconceptions?

Another prediction of the competing coexistence or dual processing account is that even after children have acquired a scientifically accurate understanding of natural selection, this understanding has to compete with persistent intuitive (and scientifically inaccurate) explanations. Thus, among children who have acquired the counterintuitive understanding of natural selection, those with greater EF resources should be less likely to display misconceptions. To test this hypothesis, we

Table 5

Logistic regression models predicting children's ability to construct a population-based understanding of natural selection as a function of their total accuracy on the mixed trials of the Flanker task (percentage), their scores on a Working Memory task (WM), and their age. OR = Odds Ratio.

	Model 1		Model 2		Model 3		Model 4	
	OR	*z scores*	*OR*	*z scores*	*OR*	*z scores*	*OR*	*z scores*
Flanker	1.07**	2.60					1.06*	2.15
WM			1.42	1.89			1.19	0.88
Age					1.03	0.58	0.99	0.23
Constant	0.005*	2.33	0.22	1.44	0.12	0.48	0.01	0.83
N	46		46		44		44	
X^2	9.80**		4.25*		0.35		9.39*	
Pseudo R^2	0.16		0.07		0.006		0.16	

* $p < .05$.

** $p < .01$.

restricted our analysis to the 28 children who from the pre-test to the comprehension-test constructed a population-based (selectionist) understanding of adaptation by natural selection after exposure to the adaptation storybook. As Table 6 shows, we then tested a series of logistic regression models in which children's tendency to display a natural selection misconception (coded as 1) on the 3-month delayed familiar animal generalization speciation post-test was regressed on their working memory score, age, and total accuracy score on the mixed trial of the flanker task. Consistent with coexistence accounts, we found that when controlling for working memory and age, children's total accuracy score on the mixed trials of the flanker task was a significant predictor of their ability to avoid displaying a misconception on the most delayed assessment they completed, OR (for a 10-percentage point increase in EF accuracy) = 0.10, $z = 2.04$, $p = .041$, Table 5.

10. Discussion

In sum, we found evidence that EF, specifically set shifting/inhibition, is involved in the construction of a population-based understanding of natural selection following children's exposure to our storybook. Restricting our analyses to children who successfully constructed such an understanding, we found evidence that once such an understanding has been constructed, EF is involved in inhibiting persistent competing intuitive explanations for that process. Thus, our results add to and extend a growing body of research in adults and children that are consistent with conceptual co-existence accounts of conceptual change: EF is needed to inhibit intuitive explanations when counterintuitive theories are constructed and are needed again when those counterintuitive theories are expressed at later time points. This is because earlier intuitive explanations are not replaced but coexist alongside later acquired counterintuitive explanations. They then compete with them during problem-solving.

11. General discussion

In older students and adults, learning natural selection represents a substantive case of conceptual change. When constructing an accurate understanding of natural selection, older students and adults do more than just acquire more knowledge. They have to overcome existing intuitively-based explanations of biological change to elaborate a qualitatively different, counterintuitive theory—one that rests on shifting focus from individuals (and their capacities for goal-directed change) to biological populations (and dynamics resulting from their inherent variation). Across two studies, our findings demonstrate that for young children, just as for adults, learning natural selection represents a process of conceptual change. Results from our pre-tests indicated that almost all second and third graders had inaccurate intuitive preconceptions about the processes of adaptation and speciation prior to instruction. Furthermore, these represented scientific misconceptions that, in many cases, implicated early developing teleological and essentialist biases that have been found to be developmentally persistent (see also Brown et al., 2020). Critically, and consistent with prior research with adults, children expressed more misconceptions at pre-test in the context of speciation than in the context of adaptation, confirming that speciation is a particularly difficult and counterintuitive concept to learn.

Despite this, over the short term, elementary school children were able to overcome their intuitions to understand and generalize an accurate population-based explanation of speciation based on a relatively circumscribed intervention that used adaptation by natural selection as the foundation for understanding speciation. The present findings not only illuminate children's capacities to benefit from this kind of mechanistic instruction and engage in this kind of challenging and complex conceptual change but also shed light on the domain-general EF mechanisms that facilitate it and the cognitive architecture that results.

Table 6

Logistic regression models predicting whether children who had constructed the theory natural selection displayed any misconception on their understanding of this process after a three-month delay as a function of their total accuracy on the mixed trial of a Flanker task (percentage), their scores on a Working Memory task (WM), and their age. OR = Odds Ratio.

	Model 1		Model 2		Model 3		Model 4	
	OR	*z scores*	*OR*	*z scores*	*OR*	*z scores*	*OR*	*z scores*
Flanker	0.79*	2.17					0.79*	2.04
WM			0.90	0.56			0.98	0.07
Age					0.95	0.87	1.00	0.03
Constant	0.00*	2.08	1.05	0.04	104.34	0.79	0.00	1.87
N	28		28		27		27	
X^2	10.27**		0.32		0.79		9.77*	
Pseudo R^2	0.28		0.01		0.02		0.27	

* $p < .05$.

** $p < .01$.

12. Educational implications

In the United States, coherent mechanistic instruction about evolution by natural selection is generally delayed until high school. Instead, even under the NGSS, elementary school children frequently receive limited instruction on piecemeal evolutionary facts (e.g., animals have specialized body parts, fossilized species show similarities to contemporary species) rather than a focus on the mechanism that connects those facts (Kelemen, 2019). This educational approach, and its underlying assumption that elementary school children are not cognitively capable of learning the complexities of natural selection, is challenged by the current results. Converging with findings from previous research, the present work yields evidence that when they are offered coherent mechanistic explanations, young children can understand complex biological and physical processes (see Kelemen, 2019). Distinctively, however, the current research also indicates that children can not only understand complex mechanistic ideas but generalize one of the most counterintuitive scientific concepts of all: the counter-essentialist idea that existing species can evolve into entirely new species. Indeed, we found that Grade 3 children who received our brief storybook intervention experienced significantly greater and more consistent learning gains on our speciation assessments than children who did not receive such instruction.

Having said this, the present research also helps clarify what may and may not be possible with particular age groups and thus how a productive developmentally-informed learning progression on evolution in elementary school might be structured. Specifically, when considered in combination with prior findings on children's learning of adaptation (Emmons et al., 2017; Kelemen et al., 2014), it suggests that it is viable to introduce 5-to-6-year-olds to adaptation by natural selection—a concept that they have been found to understand—such that by 6-to 8-years of age, children can be supported in engaging in enduring far transfer of that concept, potentially through the use of multiple analogically aligned storybooks. Based on the current results, 7- to 9-year-olds can then be scaffolded to extend their understanding of adaptation to speciation, building a foundation for even more comprehensive mechanistic understandings of why contemporary species resemble fossils and why disparate contemporary species are related. However, given evidence that the learning gains of some third graders faded after the three-month delay, yielding no significant differences in understanding between children who received the storybook intervention and those who did not, our findings reveal that an understanding of speciation should receive additional scaffolding. At minimum, exposure to another analogically-aligned storybook on speciation—an approach that has proven effective in strengthening younger children's understanding of adaptation (Emmons et al., 2017)—is likely to be helpful.

Critically, the present studies provide further support for the value of an educational approach that does not underestimate young children's explanatory learning capacities. The results show that there is value to presenting young children with coherent mechanistic explanations of counterintuitive ideas from early in development and then revisiting and elaborating these explanations over time to strengthen and enrich them (Kelemen, 2019).

12.1. Implications for counterintuitive theory construction and conceptual change

In addition to these implications for educational practice, our results also inform our theoretical understanding of the process of counterintuitive theory construction and the cognitive architecture that supports it. Increasing evidence suggests that rather than being replaced during conceptual change, intuitive explanatory frameworks coexist alongside and compete with newly-learned counterintuitive theories (e.g., Kelemen, 2019; Shtulman & Lombrozo, 2016). The present findings provide support for this picture of conceptual change in several ways. First, explorations of children's performance on the assessments showed that the storybook intervention had two complementary impacts on children's reasoning about the process of natural selection: It increased children's ability to provide a scientifically accurate population-based account of adaptation and speciation and it decreased the occurrence of misconceptions. Indeed, the latter impact was evident even after a three-month delay. One interpretation of this result is that children who received the intervention acquired a means of explaining natural selection that successfully competed with their intuitive but incorrect explanations. Second, and relatedly, children's misconceptions reasserted themselves over time. This is only possible if children's intuitive explanations continued to exist in children's minds rather than being replaced following theory construction. Finally, this competitive coexistence account of theory change is supported by our analyses of the role of EF in the construction of a population-based understanding and in the inhibition of alternative intuitive explanatory frameworks once a population-based account of natural selection has been acquired. Specifically, we found evidence that controlling for age and working memory, children's ability to learn from our intervention was predicted by higher EF scores as was their ability to inhibit misconceptions after a three-month delay (if they had acquired an accurate understanding). These results complement but also importantly extend those of Vosniadou et al. (2018) with a similar age group. Given that we controlled for working memory and used a pre- to post-intervention design, our results suggest that set shifting/inhibition are the specific aspects of EF implicated in this process of not only acquiring but maintaining and successfully mobilizing a counterintuitive theory over time.

13. Conclusion

In conclusion, young children are able to learn and abstract coherent mechanistic explanations of causally complex processes. This sheds light on children's explanatory capacities. It also has implications for early science education on pivotal counterintuitive concepts. Traditional theories of conceptual change posit the successive revision-and-replacement of explanatory theories (e.g., Carey, 1985; Gopnik & Wellman, 1994) but, in association with recent research with older children and adults (e.g., Evans et al., 2011; Kelemen & Rosset, 2009; Shtulman, 2017), the present results suggest that intuitively-based frameworks do not disappear but persistently coexist alongside formally learned theories. Thus, interventions that progressively revisit counterintuitive mechanisms from earlier in development, before misconceptions deeply entrench, promise to enhance both initial learning and longer-term online reasoning (Kelemen, 2019). Specifically, robust representations constructed during—and revisited from—earlier timepoints may reduce susceptibility to interference from coexistent intuitions later on. We look forward to additional research with larger sample sizes that extends the current replicated findings and further strengthens these conclusions.

Author contributions

S.R., D.K., and E.D. led the research design, with assistance from S.B.; S.R., S.B., and E.D. collected data; S.R. analyzed data; S.R., S.B., and D.K. wrote the paper.

Acknowledgements

We appreciate the school staff for their support, and we thank the parents and children for their participation. We would also like to thank Nishi Karim, Casey Lewry, Aimee Mather, and Emma Pitt who assisted with data collection, materials design and data coding as well as the Child Cognition Lab Research Assistants.

This research was supported by National Science Foundation Grant #1561401 & #2009176 awarded to Deborah Kelemen. The views expressed in this article are those of the authors.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2021.104635.

References

Kelemen & The Child Cognition Lab. (2017). *How the piloses evolved skinny noses*. Boston, MA: Tumblehome Learning.

Kelemen & The Child Cognition Lab. (2020). *Meet the miroungas*. Boston, MA: Tumblehome Learning.

Achieve, Inc. (2013). Next Generation Science Standards. Retrieved from http://www.nextgenscience.org.

Bascandziev, I., Powell, L. J., Harris, P. L., & Carey, S. (2016). A role for executive functions in explanatory understanding of the physical world. *Cognitive Development, 39*, 71–85. https://doi.org/10.1016/j.cogdev.2016.04.001.

Bascandziev, I., Tardiff, N., Zaitchik, D., & Carey, S. (2018). The role of domain-general cognitive resources in children's construction of a vitalist theory of biology. *Cognitive Psychology, 104*, 1–28. https://doi.org/10.1016/j.cogpsych.2018.03.002.

Bishop, B. A., & Anderson, C. W. (1990). Student conceptions of natural selection and its role in evolution. *Journal of Research in Science Teaching, 27*, 415–427. https://doi.org/10.1002/tea.3660270503.

Brown, S. A., Ronfard, S., & Kelemen, D. (2020). Teaching natural selection in early elementary classrooms: Can a storybook intervention reduce teleological misunderstandings? *Evolution: Education and Outreach, 13*, 1–19.

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

Chen, H., Cohen, P., & Chen, S. (2010). How big is a big odds ratio? Interpreting the magnitudes of odds ratios in epidemiological studies. *Communications in Statistics—simulation and Computation, 39*, 860–864.

Coley, J. D., & Tanner, K. D. (2012). Common origins of diverse misconceptions: Cognitive principles and the development of biology thinking. *CBE—Life Sciences Education, 11*, 209–215. https://doi.org/10.1187/cbe.12-06-0074.

Coley, J. D., & Tanner, K. D. (2015). Relations between intuitive biological thinking and biological misconceptions in biology majors and nonmajors. *CBE—Life Sciences Education, 14*. https://doi.org/10.1187/cbe.14-06-0094. Article 8.

Davidson, M. C., Amso, D., Anderson, L. C., & Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task-switching. *Neuropshychologia, 44*, 2037–2078. https://doi.org/10.1016/j.neuropsychologia.2006.02.006.

Diamond, A., Barnett, W. S., Thomas, J., & Munro, S. (2007). Preschool program improves cognitive control. *Science, 318*, 1387–1388. https://doi.org/10.1126/science.1151148.

Emmons, N. A., & Kelemen, D. (2015). Young children's acceptance of within-species variation: Implications for essentialism and teaching evolution. *Journal of Experimental Child Psychology, 139*, 148–160. https://doi.org/10.1016/j.jecp.2015.05.011.

Emmons, N., Lees, K., & Kelemen, D. (2017). Young children's near and far transfer of the basic theory of natural selection: An analogical storybook intervention. *Journal of Research in Science Teaching, 55*, 321–347. https://doi.org/10.1002/tea.21421.

Emmons, N. A., Smith, H., & Kelemen, D. A. (2016). Changing minds with the story of adaptation: Strategies for teaching young children about adaptation. *Early Education and Development, 27*, 1205–1221. https://doi.org/10.1080/10409289.2016.1169823.

Evans, E. M. (2001). Cognitive and contextual factors in the emergence of diverse belief systems: Creation versus evolution. *Cognitive Psychology, 42*, 217–266. https://doi.org/10.1006/cogp.2001.0749.

Evans, E. M., Legare, C., & Rosengren, K. (2011). Engaging multiple epistemologies: Implications for science education. In M. Ferrari, & R. Taylor (Eds.), *Epistemology and science education: Understanding the evolution vs. intelligent design controversy* (pp. 111–139). New York, NY: Routledge.

Evans, J., & Stanovich, K. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science, 8*, 223–241. https://doi.org/10.1177/17456916 12460685.

Geary, D. C., vanMarle, K., Chu, F. W., Hoard, M. K., & Nugent, L. (2019). Predicting age of becoming a cardinal principle knower. *Journal of Educational Psychology, 111*, 256–267. https://doi.org/10.1037/edu0000277.

Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday thought*. New York, NY: Oxford University Press.

Gelman, S. A., & Rhodes, M. (2012). Two-thousand years of stasis. In K. S. Rosengren, S. K. Brem, E. M. Evans, & G. M. Sinatra (Eds.), *Evolution challenges: Integrating research and practice in teaching and learning about evolution* (pp. 200–207). Oxford, England: Oxford University Press.

Goldberg, R. F., & Thompson-Schill, S. L. (2009). Developmental "roots" in mature biological knowledge. *Psychological Science, 20*, 480–487. https://doi.org/10.1111/j.1467-9280.2009.02320.x.

Gopnik, A., & Wellman, H. M. (1994). The theory theory. In L. A. Hirschfeld, & S. A. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (p. 257). Cambrigde, UK: Cambridge University Press.

Gregory, T. R. (2009). Understanding natural selection: Essential concepts and common misconceptions. *Evolution: Education and outreach, 2*, 156–175.

Kelemen, D. (1999). Why are rocks pointy? Children's preference for teleological explanations of the natural world. *Developmental Psychology, 35*, 1440–1452. https://doi.org/10.1037//0012-1649.35.6.1440.

Kelemen, D. (2004). Are children "intuitive theists"?: Reasoning about purpose and design in nature. *Psychological Science, 15*, 295–301. https://doi.org/10.1111/j.0956-7976.2004.00672.x.

Kelemen, D. (2012). Teleological minds: How natural intuitions about agency and purpose influence learning about evolution. In K. S. Rosengren, S. K. Brem, E. M. Evans, & G. M. Sinatra (Eds.), *Evolution challenges: Integrating research and practice in teaching and learning about evolution* (pp. 66–92). Oxford, England: Oxford University Press.

Kelemen, D. (2019). The magic of mechanism: explanation-based instruction on counterintuitive concepts in early childhood. *Perspectives on Psychological Science*. https://doi.org/10.1177/1745691619827011. Advance online publication.

Kelemen, D., Emmons, N., Seston, R., & Ganea, P. (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.

Kelemen, D., & Rosset, E. (2009). The human function compunction: Teleological explanation in adults. *Cognition, 111*, 138–143. https://doi.org/10.1016/j.cognition.2009.01.001.

Kelemen, D., Rottman, J., & Seston, R. (2013). Professional physical scientists display tenacious teleological tendencies: Purpose-based reasoning as a cognitive default. *Journal of Experimental Psychology: General, 142*, 1074–1083. https://doi.org/10.1037/a0030399.

Kelemen, D., & the Child Cognition Lab. (2018). *How the dormacks evolved longer backs*. Boston, MA: Tumblehome Learning.

Kelemen, D., Widdowson, D., Posner, T., Brown, A. L., & Casler, K. (2003). Teleo-functional constraints on preschool children's reasoning about living things. *Developmental Science, 6*, 329–345. https://doi.org/10.1111/1467-7687.00288.

Massachusetts Department of Education. (2016). Science and Technology/Engineering Learning Standards. Retrieved from http://www.doe.mass.edu/frameworks/scitech/2016-04/STE-Standards.pdf.

Metz, K. (2008). Narrowing the gulf between the practices of science and the elementary school science classroom. *The Elementary School Journal, 109*, 138–161. https://doi.org/10.1086/590523.

Shtulman, A. (2006). Qualitative differences between naïve and scientific theories of evolution. *Cognitive Psychology, 52*, 170–194. https://doi.org/10.1016/j.cogpsych.2005.10.001.

Shtulman, A. (2017). *Scienceblind: Why our intuitive theories about the world are so often wrong*. New York: Basic Books.

Shtulman, A., & Harrington, K. (2016). Tensions between science and intuition across the lifespan. *Topics in Cognitive Science, 8*, 118–137. https://doi.org/10.1111/tops.12174.

Shtulman, A., & Lombrozo, T. (2016). Bundles of contradiction: A coexistence view of conceptual change. In D. Barner, & A. S. Baron (Eds.), *Core knowledge and conceptual change* (pp. 49–67). Oxford, UK: Oxford University Press.

Shtulman, A., & Schulz, L. (2008). The relation between essentialist beliefs and evolutionary reasoning. *Cognitive Science, 32*, 1049–1062. https://doi.org/10.1080/03640210801897864.

Sinatra, G. M., Brem, S. K., & Evans, E. M. (2008). Changing minds? Implications of conceptual change for teaching and learning about biological evolution. *Evolution: Education and outreach, 1*, 189–195.

Tardiff, N., Bascandziev, I., Sandor, K., Carey, S., & Zaitchik, D. (2017). Some consequences of normal aging for generating conceptual explanations: A case study of vitalist biology. *Cognitive Psychology, 95*, 145–163. https://doi.org/10.1016/j.cogpsych.2017.04.004.

Vosniadou, S., Pnevmatikos, D., Makris, N., Lepenioti, D., Eikospentaki, K., Chountala, A., & Kyrianakis, G. (2018). The recruitment of shifting and inhibition in on-line science and mathematics tasks. *Cognitive Science, 42*, 1860–1886. https://doi.org/10.1111/cogs.12624.

Zaitchik, D., Solomon, G., Tardiff, N., & Bascandziev, I. (2016). Conceptual change: Where domain-specific learning mechanisms meet domain-general cognitive resources. In D. Barner, & A. S. Baron (Eds.), *Core knowledge and conceptual change* (pp. 73–89). Oxford, UK: Oxford University Press.