Detailed bugs or bugging details: The influence of perceptual richness changes across elementary school years

David Menendez[1]

Karl S. Rosengren[2]

Martha W. Alibali[1]

[1]University of Wisconsin-Madison

[2]University of Rochester

This is the peer reviewed version of the following article: Menendez, D., Rosengren, K. S., & Alibali, M. W. (2022). Detailed bugs or bugging details: The influence of perceptual richness changes across elementary school years. Journal of Experimental Child Psychology, *213*, 105269, which has been published in final form at 10.1016/j.jecp.2021.105269.

Acknowledgements

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Award #R305B150003 to the University of Wisconsin-Madison. The opinions expressed are those of the authors and do not represent views of the U.S. Department of Education. This research was also partially supported by a core grant to the Waisman Center at the University of Wisconsin-Madison from the National Institute of Child Health and Human Development (U54 HD090256). This research was also partially supported by a grant from the National Science Foundation (#MSN21179). Finally, this research was also supported by an undergraduate Senior Thesis Award awarded to the first author from the

Department of Psychology at the University of Wisconsin-Madison. We would like to thank Kathryn Hartfield, Erin Condon, Olympia Mathiaparanam, and Vienne Seitz for their help with data collection.

Highlights (85 character max with spaces)

Metamorphosis is a challenging concept for children to learn.

There is a developmental change on how perceptual richness influences learning.

First and second graders learned better with rich diagrams.

Fourth and fifth graders generalized more broadly with bland diagrams.

Abstract (161/250)

Visualizations are commonly used in educational materials, however not all visualizations are equally effective at promoting learning. Prior research has supported the idea that both perceptually rich and bland visualizations are beneficial for learning and generalization. We investigated whether the perceptual richness of a life cycle diagram influenced children's learning of metamorphosis, a concept that prior work suggests is difficult for people to generalize. Using identical materials, Study 1 (n = 76) examined learning and generalization of metamorphosis in first and second grade students and Study 2 (n = 53) did so in fourth and fifth grade students. Bayesian regressions revealed that first and second grade students learned more from the lesson with the perceptually rich diagram, while fourth and fifth grade students generalized more broadly with the bland diagram. The results from the fourth and fifth grade students are similar to prior research with adults. This suggests that the effect of perceptual richness on learning and generalization changes over development.

Keywords: Diagrams, Perceptual richness, Visualizations, Biological reasoning, Bayesian statistics

Detailed bugs or bugging details:

The influence of perceptual richness differs across elementary school years

Visualizations, such as graphs, diagrams, and pictures, are ubiquitous in educational materials. Visualizations are included in textbooks (Woodward, 1993), trade books (Etta & Kirkorian, 2019; Menendez, Mathiaparanam, et al., 2020), presentations (Angra & Gardner, 2018), tests (Lindner, 2020), and even classroom decorations (Fisher et al., 2014)! Given the prevalence of visualizations in educational contexts, it is important to examine which visualizations are best at promoting learning and whether they are equally effective for all students. Additionally, visualizations might be a useful tool for teaching children about difficult or counterintuitive topics, which might otherwise pose a challenge for them to learn and generalize to new instances. In this paper, we examine how the perceptual richness of a diagram (i.e., the number of visual features it contains) influences learning and transfer of a counterintuitive biological concept across the elementary school years.

Influence of perceptual richness

Many studies have examined the influence of visual representations in learning and generalization. In general, these studies find that adding visual representations to a lesson leads to better learning and generalization (Mayer, 2008, 2009; Moreno & Mayer, 1999; Wiley, 2018). However, not all visual representations are equally beneficial, as their effectiveness at promoting learning and generalization depends on characteristics of the representation (Schnotz & Bannert, 2003; Schnotz & Kürschner, 2008; Skulmowski & Rey, 2018).

One characteristic that has received a lot of attention in the psychological literature is how the information is depicted. For example, the life cycle of a ladybug can be depicted in a realistic manner with photographs or detailed drawings or in a schematic manner with line drawings (Menendez, Rosengren, et al., 2020). The literature on visualizations has not been consistent in the terminology used to describe this contrast, with realistic drawings sometimes being described as concrete, grounded, specific, perceptually rich, iconic, depictive, or as containing seductive, extraneous, or irrelevant details (Belenky & Schalk, 2014; Kaminski & Sloutsky, 2013; Koedinger et al., 2008; Menendez, Rosengren, et al., 2020; Skulmowski & Rey, 2020b). Likewise, line drawings have been described as abstract, idealized, generic, perceptually bland, symbolic, schematic, sparse, or as containing only relevant details (Butcher, 2006; Rey, 2012; Wiley, 2018; Wiley et al., 2017). Although the definitions of each of these terms are not perfectly overlapping (e.g., a diagram containing only relevant details might not be symbolic; Belenky & Schalk, 2014), there is considerable overlap in how they are instantiated in research studies. For example, abstract representations (representations that depict general concepts rather specific instantiations of those concepts) also tend to have fewer details than concrete representations. Put another way, concrete representations tend to be perceptually rich (Castro-Alonso et al., 2016). In this research, we will use the terms *perceptually rich* and *perceptually bland* as they represent our process in creating the visualizations. We define perceptual richness in terms of the number of visual features included in the representation. Additionally, the terms perceptually rich and bland are useful when reviewing the literature, as they can be applied to two-dimensional representations like photographs and diagrams, and also to three-dimensional representations like manipulatives (Carbonneau et al., 2020).

Many studies have shown that the perceptual richness leads to lower learning and transfer in adults (Butcher, 2006; Goldstone & Sakamoto, 2003; Goldstone & Son, 2005; Harp & Mayer, 1998; Mayer et al., 2008; Menendez, Rosengren, et al., 2020; Rey, 2012; Sung & Mayer, 2012) and children (Carbonneau et al., 2020; Kaminski & Sloutsky, 2013). For example, Kaminski and

Sloutsky (2013) found that teaching kindergarten to second-grade students how to read bar graphs using perceptually bland graphs led to better transfer than teaching them with perceptually rich graphs. Perceptually bland representations have been proposed to promote generalization because they make it easier for learners to discern the underlying structure of the concept (Menendez, Rosengren, et al., 2020). Perceptually rich representations might be detrimental because they contain irrelevant details that the learner needs to process, which taxes their cognitive resources while not increasing learning of the relevant material (Rey, 2012). This suggests that, for adults, and perhaps for children, rich representations can be distracting, which can inhibit learning. Additionally, rich representations can inhibit transfer as students may interpret them as overly specific.

However, some recent studies have shown that rich representations can promote learning. Several studies have suggested that rich representations can promote learning if the details they contain are not distracting and are relevant to the task at hand (Belenky & Schalk, 2014; Siler & Willows, 2014; Trninic et al., 2020). There is also support for the idea that rich representations are better for generalizing to other rich representations, as the richness might serve as a retrieval cue (De Bock et al., 2011; Skulmowski & Rey, 2020a, 2020b). Additionally, there is support for the idea that children learn and transfer better when lessons begin with rich representations and then bland representations are slowly introduced. This procedure is referred to as concreteness fading (Fyfe et al., 2014; Fyfe, McNeil, & Borjas, 2015; McNeil & Fyfe, 2012) or concrete-representational-abstract sequence (Butler et al., 2003; Flores, 2010). Taken together, the results of these studies suggest that children might benefit from rich representations when they are first learning a topic, or when the representations contain only relevant information.

The effect of visualizations on learning and transfer also depends on contextual factors. Contextual factors are features of the learning environment other than the visualization, such as the wording of the lesson (Son & Goldstone, 2009), the presence of other visualizations (Rau, 2017), whether students can collaborate (Carbonneau et al., 2020), and the instructors gestures' during the lesson (Alibali et al., 2014). One contextual factor explored in prior research is the language and labels used during the lesson. The labels used in a lesson can be specific to the exemplar being described or can be more general, conveying the idea that the information applies to a broader set of exemplars. Lessons with rich representations can promote generalization if the language is general (Flynn et al., 2020; Son & Goldstone, 2009). Regardless of the language used during the lesson, children's production of general language after the lesson has been shown to predict their generalization (Fyfe, McNeil, & Rittle-Johnson, 2015).

Additionally, the effectiveness of visualizations also depends on learner characteristics, such as prior knowledge (Goldstone & Sakamoto, 2003), working memory (Sanchez & Wiley, 2006), spatial ability (Hegarty & Sims, 1994), and interest in the domain (Cooper et al., 2018). Several studies have suggested that students with low prior knowledge benefit more from bland representations than rich ones, while students with high prior knowledge perform similarly with both types of representation (Cooper et al., 2018; Goldstone & Sakamoto, 2003; Menendez, Rosengren, et al., 2020). Taken together, this past work suggests that factors like prior knowledge and the use of abstract language could moderate the effect of perceptual richness on learning and transfer.

Understanding of life-cycle changes

The present studies focus on how children understand life cycle changes, and in particular metamorphosis. We focus on children's understanding of metamorphosis because prior research suggests that it is a difficult concept for people to grasp (Herrmann et al., 2013). People tend to believe that organisms can change in certain ways throughout their lives; for example, they may get bigger and their proportions might change. However, people typically reject more drastic changes in color and form except for organisms that are familiar (French et al., 2018; Rosengren et al., 1991). Therefore, children and adults often do not think of drastic changes, such as metamorphosis, as a viable form of biological change (French et al., 2018; Rosengren et al., 1991).

Even after instruction, students do not think that many organisms undergo metamorphosis as part of their life cycle. According to the Next Generation Science Standards (2013), which are standards for science education for students in the United States, students are expected to learn about metamorphosis in third grade. However, adults (who likely received some formal instruction on metamorphosis) generally reject metamorphosis as a possible change, both for unfamiliar species and for familiar species like ladybugs (French et al., 2018; Menendez, Rosengren, et al., 2020). Even after directly observing a caterpillar turn into a butterfly, children are often reluctant to transfer this knowledge to other animals that undergo this change (Herrmann et al., 2013; Shepardson, 2002). This is the case even though most insects and amphibians undergo metamorphosis, and thus broad generalization is often correct. This failure to generalize has been attributed to the fact that metamorphosis violates people’s expectations that animals only get bigger with age, with very few other changes to their form or shape (French et al., 2018). Thus, metamorphosis can be considered a counterintuitive topic in biology education.

One benefit of focusing on a counterintuitive concept, such as metamorphosis, is that we can use the same materials and lessons to test people of different ages. French et al. (2018) used the exact same stimuli to test 3- to 11-year-old children's and adults' intuitions about which animals undergo drastic changes such as metamorphosis. Additionally, Menendez, Rosengren and Alibali (2020) showed that adults could learn and generalize from a short lesson on metamorphosis that was designed for elementary school students. Therefore, we can examine the influence of diagrams on learning and transfer of knowledge about metamorphosis across a variety of age ranges using identical materials.

Visual representations in biology education

Given that this research focuses on how children learn a biological concept, it is also important to consider the typical visualizations used in biology education. Wiley et al. (2017) analyzed the visualizations found in middle school, high school and college biology textbooks. They found that in middle school, about half of the visualizations were perceptually rich, and the proportion of perceptually rich visualizations decreased as grade level increased. Similarly, Menendez, Johnson, et al. (2020) analyzed visualizations in elementary school textbooks, as well as trade books meant to teach elementary school children biological concepts. They found that books targeting children in early elementary school had predominantly perceptually rich visualizations, such as photographs. They also found that the proportion of rich representations decreased with grade level, such that books targeted at late elementary school students had about half bland and half rich representations. These content analyses suggest that the proportion of visualizations that are rich is highest in early elementary school, and that this proportion slowly decreases, such that most of the visualizations used in college curricular materials are bland.

Content analyses of life cycle diagrams, the type of diagram used in the present studies, also suggest there is variation in the perceptual richness of these diagrams. Menendez, Mathiaparanam, et al. (2020) analyzed life cycle diagrams found in textbooks, trade books and online. They found that the majority of the life cycle diagrams had bland backgrounds but depicted the focal animal in a rich way. However, there were some diagrams that used bland depictions of the focal animal, such as line drawings or labels.

Current studies

The current studies examine the effects of perceptual richness on children's learning and generalization of a biological concept, metamorphosis. We examined children's ability to generalize the concept of metamorphosis, as prior work shows that people have difficulty generalizing this concept beyond frogs and butterflies. Study 1 focuses on children in first and second grade. Given that the Next Generation Science Standards suggest that children should learn about metamorphosis and other life cycle changes by third grade, we tested children in first and second grade because they likely have had little exposure to or formal lessons on metamorphosis.

Our studies used a pretest-lesson-posttest design. The pretest assessed participants' knowledge of metamorphosis before the lesson. The pretest also served to replicate the findings from French et al. (2018) that children do not endorse metamorphosis as a possible change, even when it is the correct type of change for a given animal. The lesson taught children about metamorphosis in ladybugs, a familiar animal that people do not typically think undergoes metamorphosis (Menendez, Rosengren, et al., 2020). Participants received the lesson with either a perceptually rich or a perceptually bland life cycle diagram. The posttest examined whether children learned the concept in the lesson, whether they transferred their knowledge to other animals that undergo metamorphosis, and whether they overextended their knowledge to other animals that do not undergo this change.

In prior work, perceptually rich diagrams have included distracting or irrelevant information (e.g., Harp & Mayer, 1998; Mayer et al., 2008; Rey, 2012). For this reason, it is difficult to know if adding *any* information to a lesson influences learning or if only adding *irrelevant* information has an effect. To avoid this confound, the perceptually rich diagram in the current studies included only relevant details that would help learners identify the animal displayed in the diagram. The bland diagram in our study was created by removing details to the rich diagram. This makes the two diagrams more comparable, and more similar to each other, than in previous studies. Therefore, our studies provide a stringent test of the effects of adding or removing perceptual information to a lesson, even when the information is relevant.

At pretest, children were presented with a number of different animals and asked about possible changes that could occur over the lifespan. We expected children to endorse change in size more than change in color, change in color more than metamorphosis and metamorphosis more than species change, and we expected that participants would endorse metamorphosis more for animals that actually do undergo metamorphosis (French et al., 2018; Menendez, Rosengren, et al., 2020). We expected that children would endorse metamorphosis for the ladybug more at posttest than at pretest, as they had just received a lesson on the topic, and prior work shows that people endorse metamorphosis for the animal included in the lesson (Herrmann et al., 2013; Menendez, Rosengren, et al., 2020). This finding would show that children were able to learn from both lessons. However, children might learn better (that is, endorse metamorphosis more for ladybugs) if they receive the lesson with the rich diagram, as prior work shows that children learn well with rich materials (De Bock et al., 2011). Based on previous findings from Kaminski et al. (2008) and Menendez, Rosengren and Alibali (2020), we further expected that children who received the lesson with the bland diagram would transfer more (i.e., would endorse metamorphosis for non-ladybug insects) than children who received the lesson with the rich diagram. We also expected low levels of overextension, given that people do not typically endorse metamorphosis (French et al., 2018; Menendez, Rosengren, et al., 2020). Finally, we explored whether children's use of general labels when recalling the animal in the lesson and their prior knowledge would moderate the effect of perceptual richness on transfer.

Study 1

Method

Participants

We recruited 76 first (n = 38) and second grade (n = 38) children from a database of local families of children who had participated in previous studies (38 boys, 35 girls, and 3 who did not report gender). In the United States, children in first grade are typically 6- to 7-years-old, and children in second grade are typically 7- to 8-years-old. The families in this database had been recruited through local private and public schools, the local children's museum, and emails to employees at a large research university. The racial/ethnic make-up of the sample, as reported by the parents, was 58 (76.3%) White, 5 (6.6%) Asian or Asian American, 4 (5.3%) Black or African American, 2 (2.6%) Hispanic or Latinx, 1 (1.3%) Native American, 1 (1.3%) bi- or multi-racial, and 5 who did not report race or ethnicity information. Families received $15 for participating in the study.

Design overview

The study was divided into three different sections: pretest, lesson, and posttest. The pretest served as a partial replication of French et al. (2018) by examining children's endorsement of different types of changes. For each animal we asked about four different types of life cycle changes (size only, color, metamorphosis, and species) with two questions (across the lifespan and from parent to offspring). The lesson lasted 2 minutes and focused on the life cycle of a ladybug. During the lesson, children saw either a perceptually rich or perceptually bland diagram. Finally, the posttest was similar to the pretest, except that it included more animals. The posttest contained three types of items: learning items (ladybugs or similar looking beetles), transfer items (non-ladybug insects, to which generalization is appropriate), and over-extension items (non-insect animals, to which generalization is not appropriate).

Materials

All of the stimuli, diagrams, and lesson scripts can be found at https://osf.io/rqnem/?view_only=91450b4611044b3f95453db5ee6dc8f4 . The stimuli and lessons used in this study are identical to those used with adults in Menendez, Rosengren and Alibali (2020). At pretest and posttest, we asked children to accept or reject four different types of change with two different questions. This yielded eight questions per animal. We included five animals at pretest (butterfly, ladybug, grey ladybug, fish, dog) and 10 animals at posttest (ladybug, orange ladybug, firefly, stag beetle, ant, butterfly, praying mantis, fish, frog, dog). Of these animals, only the fish and the dog do not undergo metamorphosis.

In each trial, participants were presented with two images. The base form of the animal was presented on the left and the target form (i.e., the changed animal) on the right. In size change trials, the target animal was identical to the base animal, except in its size. For animals that do not go through metamorphosis, the target animal also changed in proportions to show the correct type of change (Lorenz, 1971). In color change trials, the target animal changed both in size and color. In metamorphosis trials, the target animal was the biologically correct form of the insect. For animals that do not go through metamorphosis, the "metamorphosis" trials showed a change in species. In species change trials, the target animal was of a different species. We asked children about each change with both *lifespan* questions ("When the one on the left grows up, could it look like the one on the right?") and *offspring* questions ("Could the one on the left have a baby that looks like the one on the right?"). For the lifespan question, the target form was always bigger than the base form. For the offspring questions, the target form was always smaller than the base form. The target was always different in size as prior work suggests that children do not think that changes in color and form are possible if they are not accompanied by changes in size (Rosengren et al., 1991). Samples of the base and target form for animals that do and do not go through metamorphosis for both question types can be seen in Figure 1. A sample trial can be seen in Figure 2.

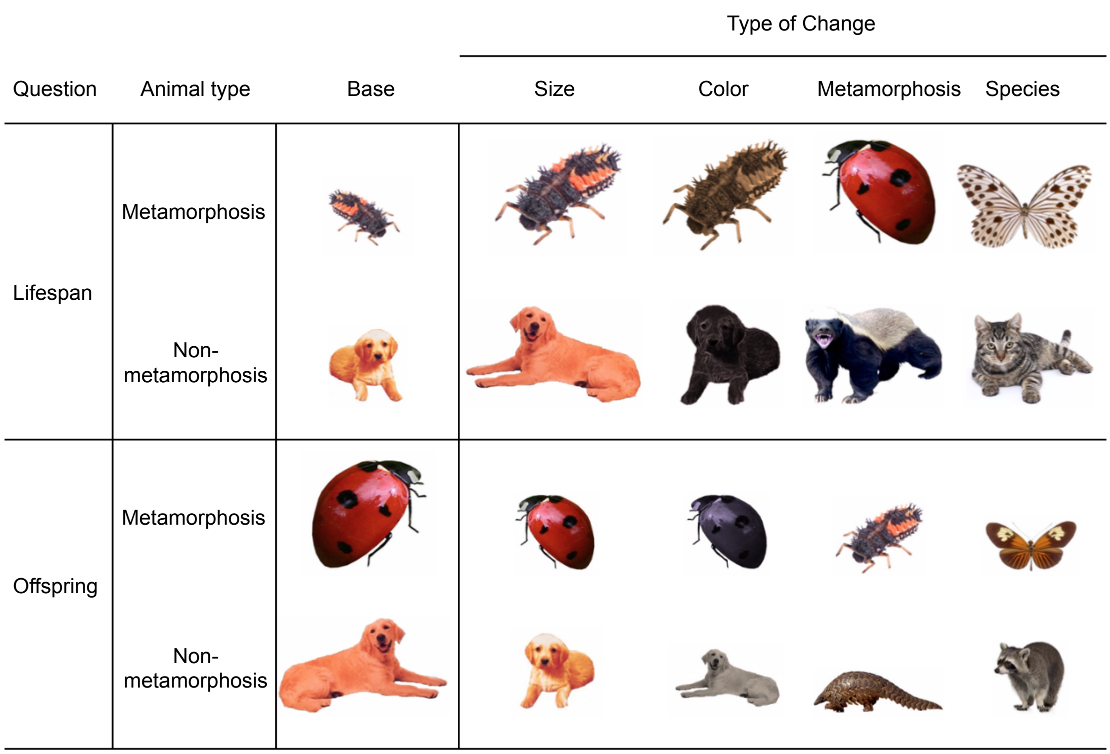

Figure 1. Sample stimuli for both question type (lifespan and offspring) and animal types (metamorphosis and non-metamorphosis). The animals were always presented in pairs. The base was always presented on the left and changed image was presented on the right.

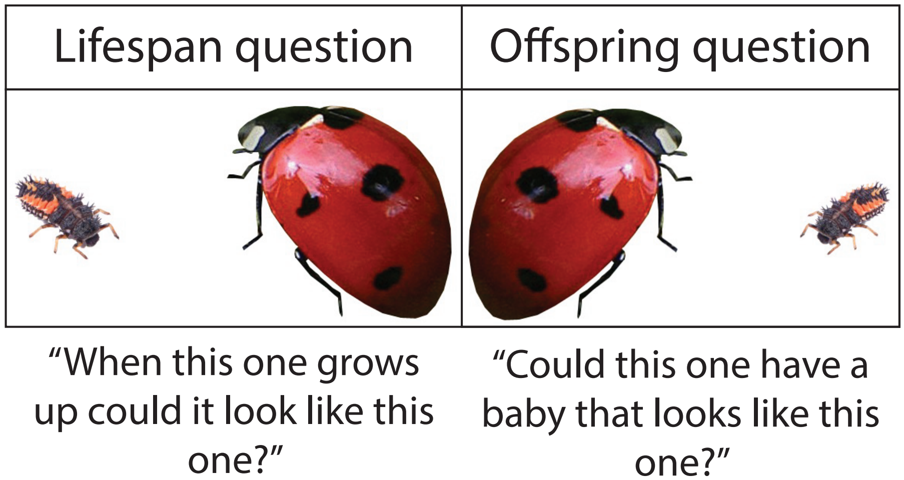

Figure 2. Sample stimuli for both types of questions.

The lesson focused on the lifecycle of a ladybug and it was delivered by a trained experimenter. The experimenter first presented the diagram and then gave the scripted, 2-minute lesson (see supplemental materials for full script). The diagram was either perceptually rich or perceptually bland, depending on the participant's condition assignment (see Figure 3). The two diagrams were identical with the exception that the perceptually rich diagram had more details including color, shading, and small features. The experimenter pointed at the image depicting each stage the first time it was mentioned. The stages mentioned were: "egg," "larva," "pupa," and "adult ladybug." The lesson noted that "many animals go through metamorphosis" but did not mention which animals do so. Therefore, we can examine how far children generalize from the lesson.

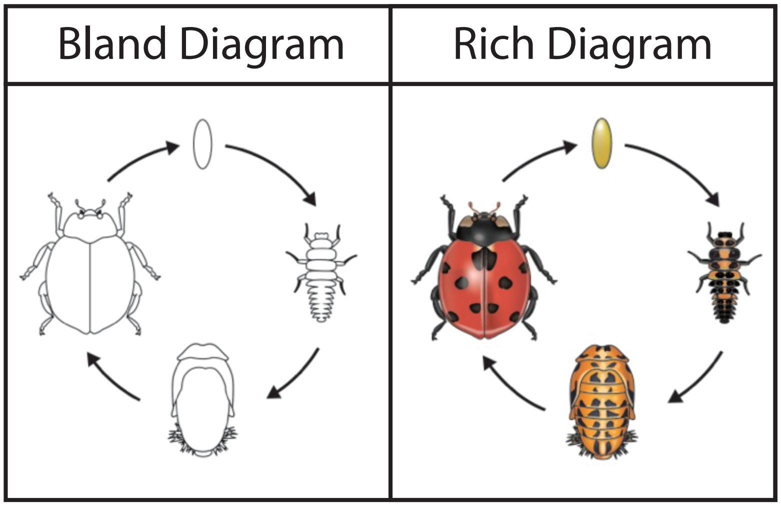

Figure 3. Bland (left panel) and rich diagram (right panel) used in the lesson. Everything else about the lesson was the same in both conditions.

Procedure

The stimuli were blocked by question, such that participants completed either all lifespan or all offspring questions first. This order was counterbalanced between participants, and the assigned order was used for both pretest and posttest. Within each question type, trials were blocked by animal, and the order of the animals was the same for all participants. The order for each trial type was randomized for each animal (but was the same for all participants). All stimuli (including the diagram) were presented on a desktop computer. The experimenter pointed at each form of the animal when asking the question.

Children first completed the pretest. Then, children received the lesson on the metamorphosis of the ladybug. After the lesson, children were asked to recall the label for each of the stages shown in the diagram while the experimenter pointed at the stage. If the children provided an incorrect label, the experimenter corrected them by providing the correct label. After the recall questions, children completed the posttest.

While the child completed the study, their parent filled out a demographic form where they could report the child's age, gender, race/ethnicity, and grade in school.

Data analysis

All of the analyses presented in this paper were done under a Bayesian framework using the *rstan* (Stan Development Team, 2020) and *rstanarm* packages (Goodrich et al., 2020) in R (R Core Team, 2020; for an overview of Bayesian data analysis, see Kruschke & Liddell, 2018). In all of the models, the priors were normal distributions. The mean of all the prior distributions was based on estimates derived from prior research by French et al. (2018) or Menendez Rosengren and Alibali (2020). For variables that were not considered in this prior work, we used a mean of 0. The standard deviation of all the prior distributions was set to make small positive and negative effects likely. This type of priors has been referred to as "weakly informative." The priors for all of the models can be found in Tables 1-4. A sensitivity analysis showing results of the same models under different priors can be found in the supplemental materials. We ran four chains with 2,000 iterations each, 1000 warm-up draws and 1000 sampling draws, yielding 4000 total draws from the posterior distribution. For each model, we report the betas (the median of the posterior distribution), 89% highest density intervals (HDI), and the probability of directions (PD; the probability that the effect is in the same direction as the beta). We report the 89% HDI, rather than the more common 95%, because the results of the 89% interval tend to be more stable (Kruschke, 2015). The 89% HDI includes the most probable (also called credible) values of the posterior distribution. Therefore, if zero is not included in the in the HDI, it means that zero is an unlikely value for the effect of the variable.

We also coded children's responses to the recall questions that were posed right after the lesson. Following prior research by Menendez, Rosengren and Alibali (2020), we coded participants' responses to the last state as either *general* or *specific*. General labels were responses that did not mention the category of ladybug, such as "adult," "beetle," "insect," or "adult stage." Specific labels were responses that mentioned the category of ladybug, such as "ladybug," or "adult ladybug." All of these labels are correct, but they differ in whether children are remembering the information as specific to the ladybug or remembering it as more broadly applying to other animals. The labels used in the lesson were specific ("adult ladybug") therefore general labels, if provided, were spontaneously generated by the child.

Results

Pretest performance

We analyzed the *proportion of trials* on which children endorsed each of the four types of change (size, color, metamorphosis, species), for different animal types (metamorphosis, non-metamorphosis) and different question types (lifespan, offspring). We used dummy codes to examine the effect of type of change, and we set change in size as the reference category. We also included the interaction of these three factors and all lower-level interactions, and we controlled for the effect of grade (first or second). We included by-subject random intercepts and by-subject random slopes for type of change, animal type, and question type (and the respective interactions). Due to convergence issues, we did not allow the random effects to correlate. The means reported are unadjusted mean proportions of endorsements.

As hypothesized, children endorsed change in size (M = 0.68, SD = 0.32) more than change in color (M = 0.35, SD = 0.34), b = -0.33 [-0.38, -0.28], color more than metamorphosis (M = 0.29, SD = 0.32), b = -0.06 [-0.09, -0.02], and metamorphosis more than species change (M = 0.06, SD = 0.16), b = -0.24 [-0.26, -0.21]. However, the pattern differed for metamorphosis and non-metamorphosis animals, as shown by interactions between animal type and the size change and metamorphosis contrast, *b* = -0.45 [-0.53, -0.37], animal type and the color change and metamorphosis contrast, *b* = -0.45 [-0.51, -0.40], and animal type and the species change and metamorphosis contrast, *b* = -0.34 [-0.39, -0.29]. In order to explore these interactions, we recentered our model at each type of change, and looked at the simple effect of animal type. Children endorsed metamorphosis, *b* = 0.39 [0.35, 0.43], and species change more, *b* = 0.05 [0.02, 0.07], and change in size, *b* = -0.06 [-0.10, -0.02], and color less, *b* = -0.06 [-0.10, -0.02] for animals that go through metamorphosis than for animals than do not go through metamorphosis. Additionally, children were more likely to endorse species change for lifespan questions for animals that undergo metamorphosis, *b* = 0.18 [0.12, 0.23]. See Figure 4. There was no evidence for an effect of grade, *b* = -0.02 [-0.06, 0.02].

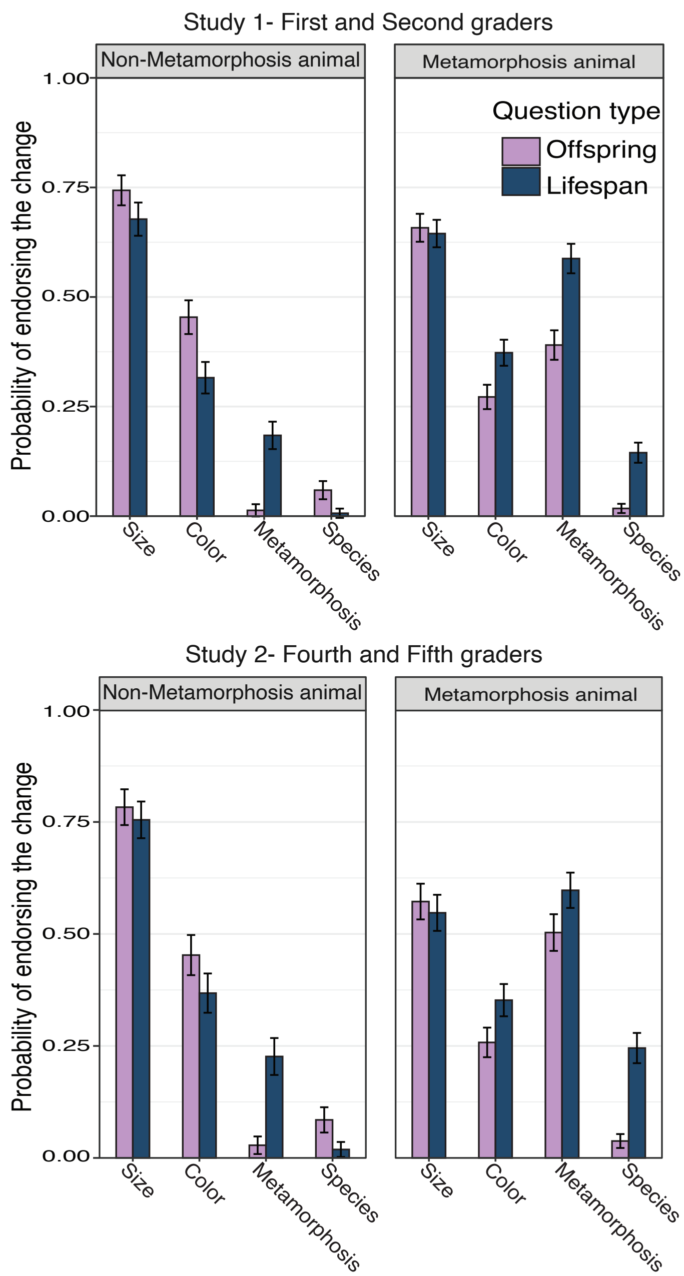

Figure 4. Probability of endorsing each type of change, broken down by question type. The left panels show endorsements for animals that do not go through metamorphosis and the right panels show endorsements for animals that go through metamorphosis. The top panels show the results for Study 1 and the bottom panels show the results for Study 2. The error bars represent within-subject standard errors.

Lesson

We predicted the number of stages children correctly labelled right after the lesson from diagram condition (rich versus bland), pretest score, grade, the diagram condition by pretest interaction, and the diagram condition by grade interaction. Children in the rich condition (*M* = 3.35, *SD* = 0.79) and the bland condition (*M* = 3.26, *SD* = 0.86) correctly labelled similar numbers of stages. See Figure 5. We found evidence for an effect of grade, such that second graders (*M* = 3.68, *SD* = 0.53) correctly labelled more stages than first graders (*M* = 2.95, *SD* = 0.90). We did not find evidence for any other effects or interactions. The model summary can be seen in Table 1.

We also examined whether the label that children provided for the final stage was specific (e.g., "ladybug") or general (e.g., "insect" or "adult"). We predicted the probability of children providing a general label from diagram condition, pretest score, their interaction, grade, and the grade by diagram interaction. We found evidence that children in second grade (*M* = 0.54, *SD* = 0.51) were more likely to provide a general label than were children in first grade (*M* = 0.11, *SD* = 0.31), *OR* = 10.83, *b* = 2.38 [1.29, 3.37]. We found no evidence for an effect of diagram condition, *OR* = 0.55, *b* = -0.60 [-1.65, 0.43], or pretest, *OR* = 0.99, *b* = -0.01 [-0.27, 0.23], nor an interaction between pretest and diagram condition, *OR* = 0.76, *b* = -0.27 [-0.77, 0.20], or between diagram condition and grade, *OR* = 0.54, *b* = -0.63 [-2.53, 1.23].

Table 1. Model summary for Study 1 and 2 for the number of labels correctly recalled. The table reports the priors, mean of the posterior distribution for each parameter (b), the 89% highest density interval (HDI), and the probability of direction (PD).

	Study 1				Study 2			
Variable	Priors	b	89% HDI	PD	Priors	b	89% HDI	PD
Diagram condition	N(0.17, 2.5)	0.14	-0.13, 0.43	78.97%	N(0.14, 2.5)	**-0.25**	**-0.50, -0.01**	**94.90%**
Pretest score	N(0.11, 2.5)	0.03	-0.04, 0.11	77.53%	N(0.03, 2.5)	-0.03	-0.10, 0.04	74.08%
Grade	N(0, 2.5)	**0.71**	**0.42, 0.98**	**100.00%**	N(0.71, 2.5)	0.24	-0.03, 0.48	93.30%
Diagram x Pretest	N(0, 2.5)	-0.1	-0.24, 0.04	87.28%	N(-0.1, 2.5)	-0.12	-0.28, 0.01	91.90%
Diagram x Grade	N(0, 2.5)	-0.08	-0.61, 0.53	58.38%	N(-0.08, 2.5)	0.27	-0.24, 0.77	79.40%

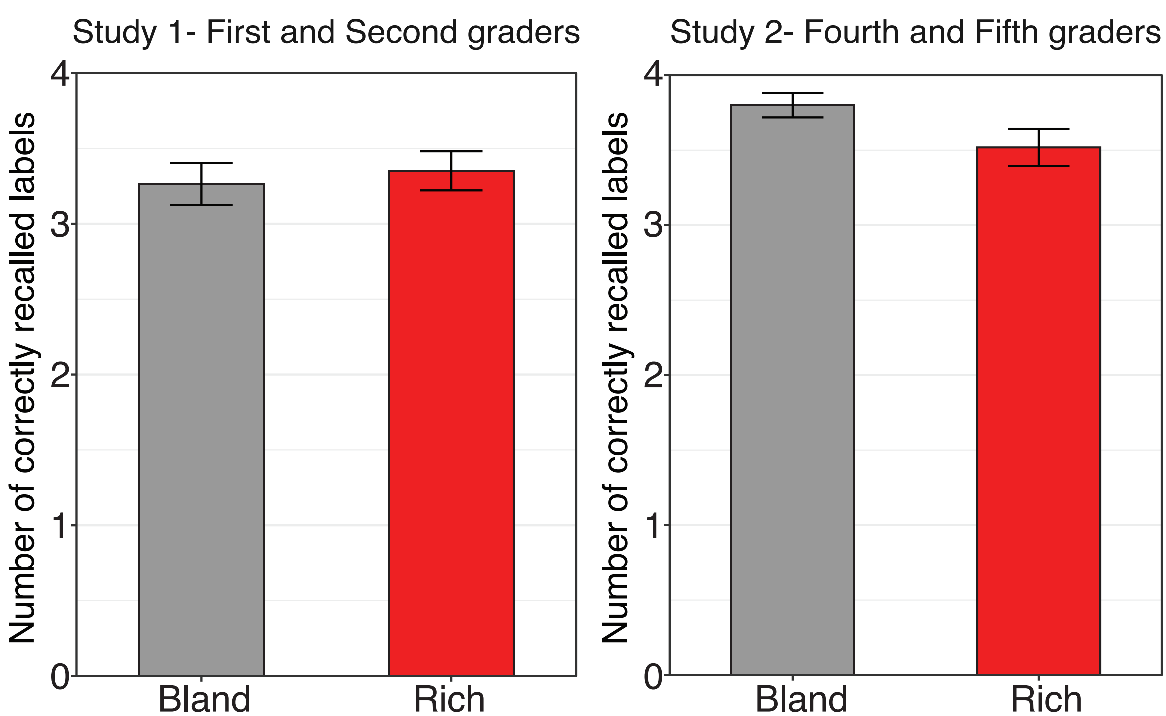

Figure 5. Average number of labels correctly recalled by children who received the lesson with the bland (left-most bars) or rich (right-most bars) life cycle diagram. The left panel shows the

results for Study 1 and the right panel shows the results for Study 2. The error bars show between-subject standard errors.

Learning

To examine if children were more likely to endorse metamorphosis after the lesson, we compared children's responses to the ladybug items at pretest and posttest. We fit a generalized linear mixed-effects model with a binomial link function predicting children's endorsement of metamorphosis for the ladybug items from test time (pretest versus posttest), diagram condition (rich versus bland), their interaction, number of correct labels provided after the lesson, grade, and question type (offspring versus lifespan). We also included by-subject random intercepts and by-subject random slopes for test time and question type. The model did not converge when we allowed the random effects to correlate, so we removed the correlations of the random effects.

As predicted, children were more likely to endorse metamorphosis at posttest ($M = 0.66$, $SD = 0.47$) than at pretest ($M = 0.29$, $SD = 0.45$). There was no effect of diagram condition, but there was a test time by diagram condition interaction. As can be seen in Figure 6, at posttest children who received the lesson with the rich diagram were more likely to endorse metamorphosis ($M = 0.74$, $SD = 0.44$) than children who received the lesson with the bland diagram ($M = 0.58$, $SD = 0.50$). Children were also more likely to endorse metamorphosis for the lifespan questions ($M = 0.52$, $SD = 0.50$) than the offspring questions ($M = 0.43$, $SD = 0.50$). Additionally, second graders ($M = 0.56$, $SD = 0.50$) were more likely to endorse metamorphosis than first graders ($M = 0.39$, $SD = 0.49$). We found no evidence for an effect of the number of labels correctly recalled. Model summaries can be seen in Table 2.

Table 2. Model summary for Study 1 and 2 for the endorsement of metamorphosis for ladybug items. The table reports the priors, mean of the posterior distribution for each parameter (b), the 89% highest density interval (HDI), and the probability of direction (PD).

	Study 1					Study 2				
Variable	Priors	b	OR	89% HDI	PD	Priors	b	OR	89% HDI	PD
Test time	N(25.46, 2.5)	**3.74**	**42.25**	**2.90, 4.60**	**100.00%**	N(3.74, 2.5)	**3.94**	**51.37**	**2.99, 4.73**	**100.00%**
Diagram condition	N(0.6, 2.5)	0.56	1.75	-0.47, 1.59	81.62%	N(0.56, 2.5)	-0.72	0.49	-1.62, 0.19	90.33%
Question type	N(4.68, 2.5)	**0.95**	**2.58**	**0.50, 1.38**	**99.98%**	N(0.95, 2.5)	0.32	1.38	-0.15, 0.81	86.15%
# of labels	N(0.24, 2.5)	0.51	1.66	-0.19, 1.25	87.52%	N(0.51, 2.5)	0.45	1.57	-0.36, 1.21	82.45%
Grade	N(0, 2.5)	**1.36**	**3.88**	**0.18, 2.46**	**97.20%**	N(1.36, 2.5)	-0.25	0.78	-1.18, 0.63	66.83%
Test time x diagram type	N(0.99, 2.5)	**1.67**	**5.29**	**0.39, 2.89**	**98.38%**	N(1.67, 2.5)	0.11	1.12	-1.22, 1.57	54.83%

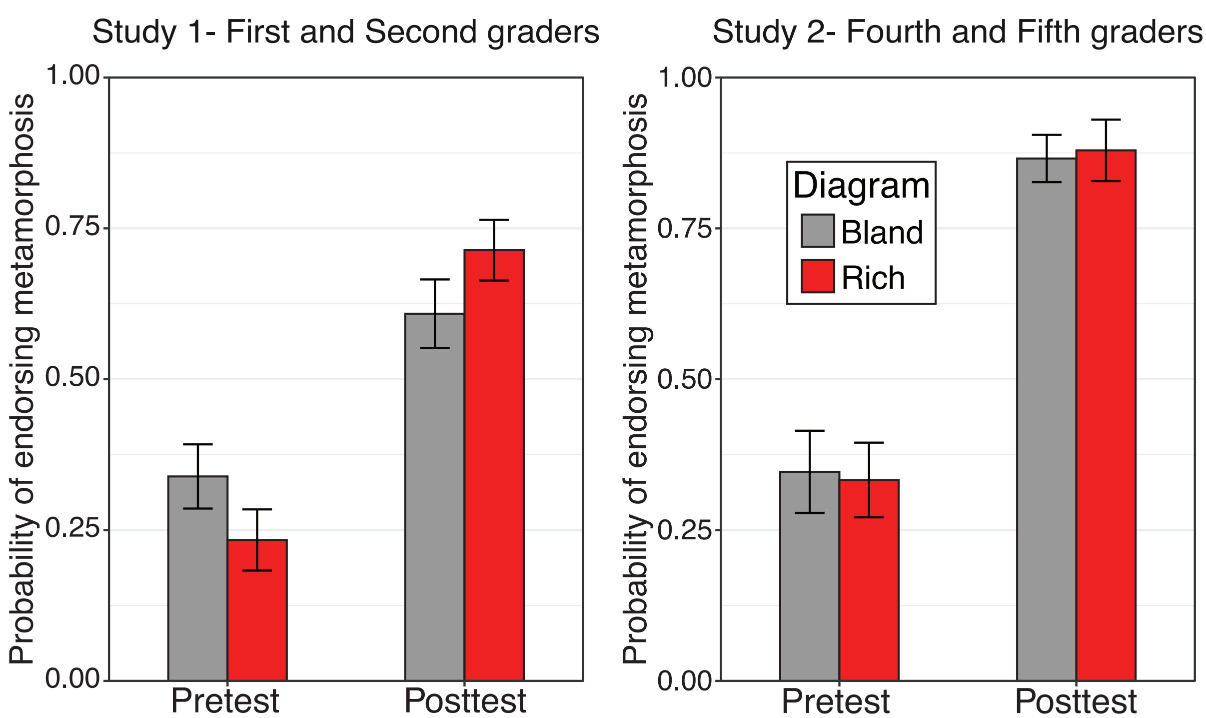

Figure 6. Probability of endorsing metamorphosis at pretest (left-most bars) and posttest (right-most bars), for children who received the lesson with the bland (grey bars) or rich (red bars) life cycle diagram. The left panel shows the results for Study 1 and the right panel shows the results for Study 2. The error bars show the within-subject standard errors.

Transfer

To examine children's generalization, we fit a generalized linear mixed-effects model predicting children's endorsement of metamorphosis for non-ladybug insects. We included diagram condition, question type, pretest, grade, and the interaction between pretest and diagram condition. Given that how much children learn is an important predictor of how much they generalize, we also included how many times they endorsed metamorphosis for the ladybug (learning items, Range = 0-4). Additionally, prior research suggests that children's use of general language predicts their generalization, so we included whether children provided a general label for the adult stage. Finally, Yzerbyt, Muller, and Judd (2004) suggest that when testing the interaction between a manipulated variable (e.g., richness of the diagram) and a measured variable (e.g., pretest scores) when controlling for a covariate (e.g., amount of learning), the estimate for the interaction is unbiased only when the model includes the covariate by manipulated variable interaction. Following this recommendation, we included interactions between diagram condition and learning scores, grade, and general labels. We also included by-subject random intercepts, by-subject random slopes for the effect of question type, and their correlation.

Contrary to our prediction, we did not find evidence that diagram condition was related to transfer. It is worth noting that the bulk (80%) of the posterior distribution is above 0. This is in the opposite direction as our priors, and contrary to previous findings with adults (Menendez et al., 2020) and to our hypothesis. Values above 0 would suggest that the tendency was for children who saw the rich diagram (*M* = 0.62, *SD* = 0.49) to transfer more than children who saw the bland diagram (*M* = 0.56, *SD* = 0.50). See Figure 7. We did find evidence that as children's pretest score increased, they were more likely to transfer. We also found evidence that children in second grade (*M* = 0.67, *SD* = 0.47) were more likely to transfer than children in first grade (*M* = 0.52, *SD* = 0.50). Additionally, children who endorsed metamorphosis more for the learning items were more likely to endorse metamorphosis for the transfer items. There were no other main effects or interactions. See model summaries in Table 3.

Table 3. Model summary for Study 1 and 2 for the endorsement of metamorphosis for non-ladybug insect items (transfer items). The table reports the priors, mean of the posterior distribution for each parameter (b), the 89% highest density interval (HDI), and the probability of direction (PD).

	Study 1					Study 2				
Variable	Priors	b	OR	89% HDI	PD	Priors	b	OR	89% HDI	PD
Diagram condition	N(-0.5, 2.5)	0.29	1.39	-0.31, 0.81	80.17%	N(0.29, 2.5)	-0.4	0.67	-0.87, 0.10	90.97%
Pretest score	N(0.15, 2.5)	**0.33**	**1.4**	**0.19, 0.48**	**100.00%**	N(0.33, 2.5)	-0.04	0.96	-0.18, 0.11	66.97%
Question type	N(-0.14, 2.5)	0.22	1.24	-0.35, 0.79	72.97%	N(0.22, 2.5)	-0.48	0.62	-1.02, 0.00	93.85%
Grade	N(0, 2.5)	**0.33**	**1.39**	**0.03, 0.63**	**95.80%**	N(0.33, 2.5)	0.36	1.43	-0.05, 0.71	93.58%
Learning score	N(0.57, 2.5)	**0.33**	**1.4**	**0.13, 0.53**	**99.67%**	N(0.33, 2.5)	**0.34**	**1.4**	**0.01, 0.67**	**94.83%**
Use of general label	N(0.43, 2.5)	0.44	1.56	-0.14, 1.09	87.72%	N(0.44, 2.5)	0.21	1.23	-0.30, 0.78	72.75%
Diagram x Pretest	N(0.26, 2.5)	-0.13	0.88	-0.42, 0.15	76.88%	N(-0.13, 2.5)	0.28	1.32	-0.02, 0.57	93.47%

Diagram x Grade	N(0, 2.5)	0.4	1.49	-0.81, 1.46	71.05%	N(0.40, 2.5)	**1.4**	**4.06**	**0.41, 2.38**	**98.92%**
Diagram x Learning	N(0.23, 2.5)	0.28	1.32	-0.13, 0.71	86.33%	N(0.28, 2.5)	-0.33	0.72	-0.99, 0.36	78.42%
Diagram x General label	N(0.28, 2.5)	0.39	1.48	-0.69, 1.65	70.65%	N(0.39, 2.5)	-0.7	0.49	-1.73, 0.45	84.90%

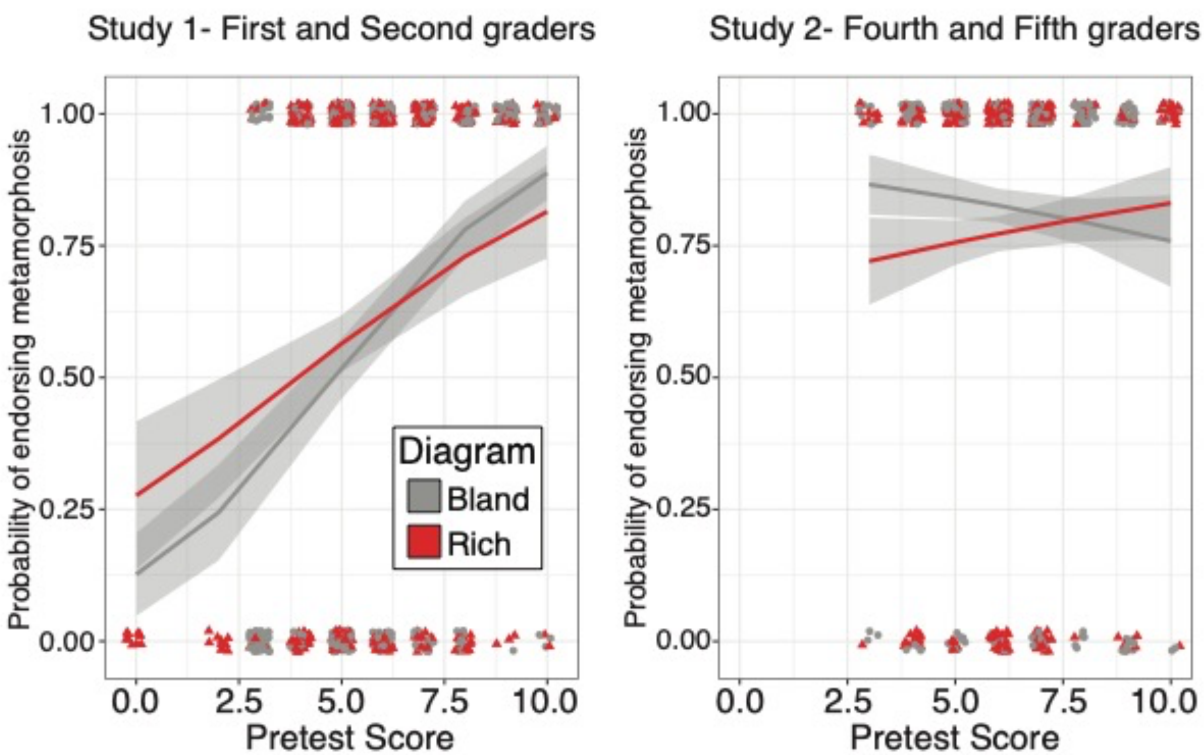

Figure 7. Probability of endorsing metamorphosis for non-ladybug insects (transfer items) broken down by pretest score (on the x-axis). The grey line shows participants who received the lesson with the bland diagram and the red line shows participants who received the lesson with the rich diagram. The left panel shows the results for Study 1, and the right panel shows the results for Study 2. The error bands show the within-subject standard errors.

Overextension

We also wished to examine whether children overextended from the lesson and endorsed metamorphosis for animals that do not undergo this change, such as dogs and fish. For these animals, the metamorphosis and species change trials are both non-biological species changes, so we combined them when looking at overextension. As expected, children rarely endorsed drastic life cycle changes for the dog (*M* = 0.04 out of 4, *SD* = 0.20), but some children did endorse these changes for the fish (*M* = 0.68 out of 4, *SD* = 0.85). Therefore, we focused on the fish items for the overextension analysis. We fit a generalized linear mixed-effects mode model with a binomial link function predicting the probability that children endorsed metamorphosis for the fish. We included test time (pretest versus posttest), diagram condition, their interaction, question type, and grade as predictors. We also included by-subject intercepts and by-subject random slopes for the effect of test time and question type, and allowed them to correlate. Model summaries are presented in Table 4.

We did not find evidence for an effect of time, diagram condition, or an interaction. There was also no effect of age. See Figure 8. However, we did find that children were more likely to overextend for the lifespan questions (*M* = 0.22, *SD* = 0.41) than the offspring questions (*M* = 0.09, *SD* = 0.28).

Table 4. Model summary for Study 1 and 2 for the endorsement of metamorphosis for fish items. The table reports the priors, mean of the posterior distribution for each parameter (b), the 89% highest density interval (HDI), and the probability of direction (PD).

	Study 1					Study 2				
Variable	Priors	b	OR	89% HDI	PD	Priors	b	OR	89% HDI	PD
Test time	N(-0.08, 2.5)	0.38	1.46	-0.03, 0.78	92.95%	N(0.38, 2.5)	0.27	1.31	-0.18, 0.74	82.73%
Diagram condition	N(-0.05, 2.5)	0.41	1.50	-0.16, 0.96	87.98%	N(0.41, 2.5)	-0.01	0.99	-0.54, 0.57	50.45%
Question type	N(0.6, 2.5)	**1.23**	**3.44**	**0.82, 1.67**	**100.00%**	N(1.23, 2.5)	**0.49**	**1.63**	**0.05, 0.92**	**96.40%**
Grade	N(0, 2.5)	0.12	1.13	-0.46, 0.65	62.98%	N(0.12, 2.5)	0.11	1.12	-0.43, 0.68	63.15%
Test time x Diagram	N(-0.31, 2.5)	-0.32	0.72	-1.10, 0.46	73.02%	N(-0.32, 2.5)	**-1.35**	**0.26**	**-2.25, -0.44**	**99.20%**

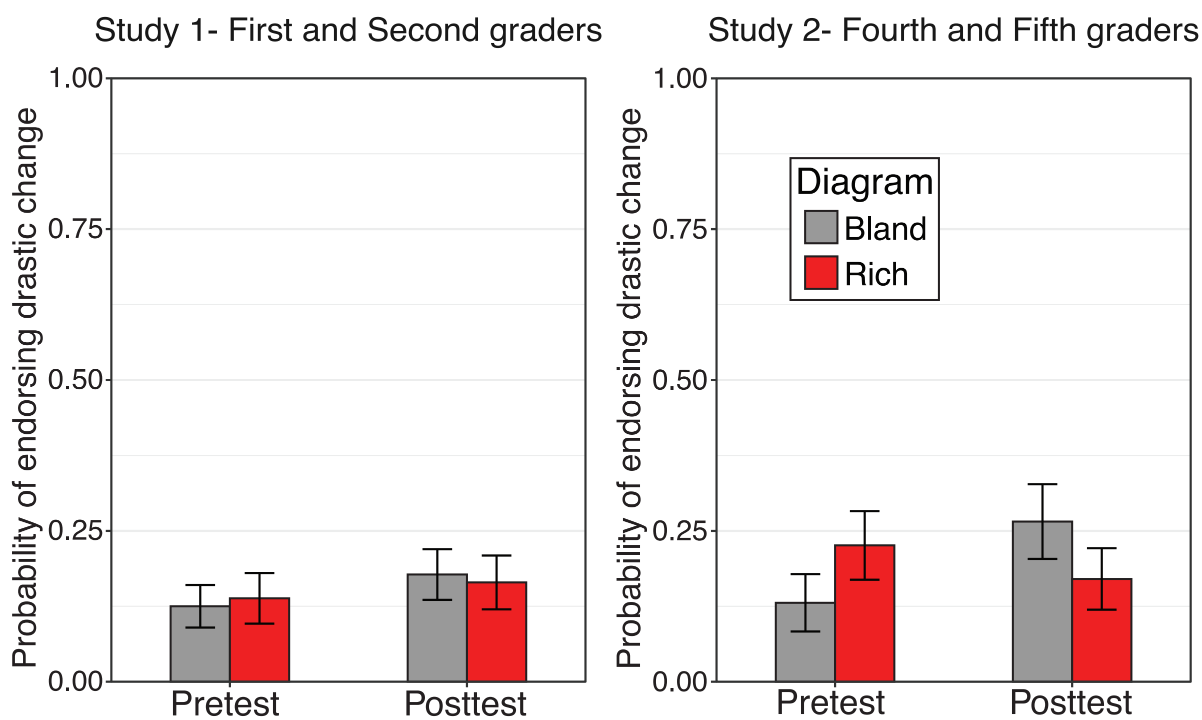

Figure 8. Probability of endorsing drastic life cycle changes for the fish items at pretest (left-most bars) and posttest (right-most bars) for children who received the lesson with bland (grey

bars) or rich (red bars) life cycle diagram. The left panel shows the results for Study 1, and the right panel shows the results for Study 2. The error bars show the within-subject standard errors.

Discussion

We examined whether the perceptual richness of diagrams influenced children's learning and generalization about metamorphosis. Overall, we found that children learned better if they received the lesson with the rich diagram than if they received the lesson with the bland diagram. We did not find a reliable effect of diagram type on generalization, but the bulk of the posterior distribution suggests that children were more likely to generalize from lessons that included the rich diagram, which is contrary to the effects found in previous work with adults (Menendez et al., 2020). This suggests that the effects of perceptual richness on children's learning and generalization are different from those for adults.

Given this surprising result, we decided to examine whether older children would show effects more similar to those found in adults. To examine how the effects of perceptual information on learning and generalization change over development, in Study 2 we tested fourth and fifth grade children. We used the same lessons and testing materials as in Study 1 and in previous research with adults (Menendez et al., 2020). We tested fourth and fifth graders because according to the Next Generation Science Standards, students should learn about metamorphosis in third grade. Therefore, the students should all have had relatively recent exposure to the concept of metamorphosis. Additionally, in these later school years, educational materials start to include more bland representations (Menendez, Johnson, et al., 2020). Therefore, we expected that fourth and fifth grade students might benefit from the bland diagram. All other predictions were the same as in Study 1.

Study 2

Method

Participants

We recruited 53 fourth (n = 30) and fifth grade (n = 23) children from the same database used in Study 1 (27 boys, 26 girls). In the United States, children in fourth grade are typically 9- to 10-years-old, and children in fifth grade are typically 10- to 11-years-old. We initially intended to collect the same number of participants as in Study 1, but we had to stop data collection due to the COVID-19 pandemic. The racial/ethnic make-up of the sample, as reported by the parents was 41 (77.4%) White, 4 (7.5%) Asian or Asian American, 3 (5.7%) Black or African American, 4 (7.5%) bi- or multi-racial, and 1 who reported another race/ethnic category. Families received $15 for participating in the study.

Materials and Procedure

The design, materials, and procedures were identical to Study 1. At the end of the study, we added two questions that asked children about their beliefs about the origin of species (adapted from Evans, 2001). These questions were added to pilot test them for a future study. These questions were “How do you think the first spider got here to Earth?” and “How do you think the first butterfly got here to Earth?” Given that beliefs about common ancestry are not central to the research questions addressed in this paper, we do not discuss these questions here. All data, materials and analyses scripts can be found in https://osf.io/rqnem/?view_only=91450b4611044b3f95453db5ee6dc8f4.

Data analysis

We used the same data analytic approach as for Study 1. We used the mean of the posterior distributions of Study 1 as our priors for Study 2. The priors for all the models can be found in Tables 1-4. A sensitivity analysis showing results of the same models under different priors can be found the supplemental materials. We ran four chains with 2,000 iterations each,1000 warm-up draws and 1000 sampling draws, yielding 4000 total draws from the posterior distribution. We also coded children's responses to the recall question about the final (adult) stage as either general or specific, in the same way as we did in Study 1.

Results

We followed the same data analysis plan as in Study 1. We first present the results for children's endorsement of life cycle change before the lesson. These analyses are a partial replication of the findings of French et al. (2018), and they show that even these children who have received formal lessons on metamorphosis in school do not always endorse metamorphosis for animals that undergo this change. Then we present the results on how perceptual richness influences children's recall from the lesson, learning, transfer and overextension.

Pretest performance

We analyzed the proportion of trials children endorsed for each type of change (size, color, metamorphosis, species), animal type (metamorphosis, non-metamorphosis), and question type (lifespan, offspring). We used dummy codes to examine the effect of type of change, and we set change in size as the reference category. We also included the interaction of these three factors and all lower-level interactions, and we controlled for the effect of grade (fourth, fifth). We included by-subject random intercepts and by-subject random slopes for type of change, animal type, and question type, and the respective interactions. Due to convergence issues, we did not allow the random effects to correlate. The means reported are unadjusted mean proportion of endorsements.

As in Study 1, children endorsed change in size (*M* = 0.66, *SD* = 0.32) more than change in color (*M* = 0.36, *SD* = 0.36), *b* = -0.31 [-0.36, -0.25] and metamorphosis (*M* = 0.34, *SD* = 0.33), *b* = -0.33 [-0.37, -0.28]. However, in this study, there was no difference in endorsement of change in color and metamorphosis, *b* = 0.02 [-0.04, 0.08]. Children also endorsed metamorphosis more than species change (*M* = 0.10, *SD* = 0.22), *b* = -0.24 [-0.27, -0.21]. As before, the pattern was different for metamorphosis and non-metamorphosis animals, as shown by interactions between animal type and the size change and metamorphosis contrast, *b* = -0.63 [-0.72, -0.55], animal type and the color change and metamorphosis contrast, *b* = -0.53 [-0.60, -0.45], and animal type and the species change and metamorphosis contrast, *b* = -0.33 [-0.40, -0.27]. In order to explore these interactions, we recentered our model for each type of change, and looked at the simple effect of animal type. Children endorsed metamorphosis, *b* = 0.42 [0.38, 0.47] and species change more, *b* = 0.09 [0.05, 0.13], and change in size, *b* = -0.21 [-0.26, -0.16], and color less, *b* = -0.11 [-0.16, -0.06] for animals that go through metamorphosis than for animals than do not go through metamorphosis. Additionally, as in Study 1, children were more likely to endorse species change for the lifespan questions for animals that undergo metamorphosis, *b* = 0.27 [0.19, 0.35]. See Figure 4. There was no evidence for an effect of grade, *b* = 0.01 [-0.04, 0.05].

Lesson

We examined whether diagram condition influenced how many stages children correctly labelled after the lesson. We included diagram condition, pretest score, grade, the diagram condition by pretest interaction and the diagram condition by grade interaction. Unlike Study 1, we found evidence that children who received the lesson with the bland diagram (*M* = 3.80, *SD* = 0.41) correctly labelled more stages than children who received the lesson with the rich diagram

(*M* = 3.52, *SD* = 0.64). There was also some evidence for a grade effect, such that fifth graders (*M* = 3.78, *SD* = 0.42) correctly labelled more stages than fourth graders (*M* = 3.55, *SD* = 0.63), as 93% of the posterior distribution was above 0. We did not find evidence for any other effects or interactions. See Figure 5 and model summaries in Table 1.

We also sought to predict whether children used general labels (e.g., "insect" or "adult") versus specific labels (e.g., "ladybug") to describe the final stage. We fit a logistic regression with diagram condition, pretest score, grade, the diagram condition by pretest interaction, and the diagram condition by grade interaction as predictors. We did not find an effect of diagram condition, *OR* = 1.20, *b* = 0.18 [-0.85, 1.29], an effect of grade, *OR* = 1.53, *b* = 0.42 [-0.52, 1.52], or an interaction between diagram condition and grade, *OR* = 5.53, *b* = 1.71 [-0.21, 3.53]. This interaction was not hypothesized, however, the bulk of the distribution (about 93%) was above 0, suggesting that fourth graders who saw the bland diagram (*M* = 0.54, *SD* = 0.52) were more likely to provide a general label than those who saw the rich diagram (*M* = 0.25, *SD* = 0.45), while the opposite was true for fifth graders (M_{bland} = 0.25, SD_{bland} = 0.45) (M_{rich} = 0.55, SD_{rich} = 0.52). We also found that as pretest scores increased, children were more likely to provide a general label, *OR* = 1.67, *b* = 0.51 [0.17, 0.87]. There was no interaction between diagram condition and pretest score, *OR* = 0.82, *b* = -0.20 [-0.95, 0.46].

Learning

We also examined whether children were more likely to endorse metamorphosis for the ladybug items after the lesson. To do so, we fitted a generalized linear mixed-effects model with a binomial link function predicting children's endorsement of metamorphosis for the ladybug items. We included test time (pretest versus posttest), diagram condition, question type, number of stages correctly labelled, grade, and the test time by diagram condition interaction. We also included by-subject random intercepts and by-subject random slopes for the effects of test time and question type. Due to convergence issues, we did not allow the random effects to correlate.

As in Study 1, we saw that children were more likely to endorse metamorphosis for ladybugs at posttest (M = 0.87, SD = 0.33) than at pretest (M = 0.34, SD = 0.47). There was no evidence for an effect of diagram condition nor an interaction between test time and diagram condition, unlike in Study 1. See Figure 6. We did not find evidence for any other effects. Model summaries can be found in Table 2.

Transfer

We next examined children's endorsement of metamorphosis for the non-ladybug insect items. We fit a generalized linear mixed-effects model with a binomial link function. We included diagram condition, pretest score, question type, grade, learning score, and whether children used general labels during the recall task as predictors. We also included interactions between diagram condition and pretest score, diagram condition and grade, diagram condition and learning score, and diagram condition and general labels. We included by-subject random intercepts and by-subject random slopes for the effect of question type. Model summaries are presented in Table 3.

We did not find a main effect of diagram condition. However, unlike in Study 1, in which the bulk of the posterior distribution were values above 0, 90% of the posterior distribution were values below 0, suggesting that children generalized more with the bland diagram (M = 0.80, SD = 0.40) than with the rich diagram (M = 0.76, SD = 0.43). There was no effect of pretest, but there was some indication of an interaction between diagram condition and pretest (93% of the posterior distribution above 0). This interaction suggests that children with low pretest scores were more likely to transfer if they saw the bland rather than rich diagram, while diagram condition did not matter for children with high pretest scores. This result is in line with our prediction and is similar to the results found with adults. See Figure 7. There was also some evidence for an effect of grade, such that fourth graders were more likely to generalize than fifth graders, with 93% of the posterior distribution below 0. We also found a grade by diagram condition interaction that was not hypothesized. We found that fourth graders transferred more with the bland diagram than the rich diagram, but fifth graders transferred equally well with both diagrams. We found similar evidence that children were more likely to endorse metamorphosis for lifespan questions (M = 0.80, SD = 0.40) than offspring questions (M = 0.75, SD = 0.43, 93% of the posterior distribution above 0). We also found that children who endorsed metamorphosis more on the learning items (i.e., the ladybug items) were also more likely to endorse metamorphosis for the transfer items. We did not find evidence for any other effects or interactions.

Overextension

As in Study 1, we also examined whether children overextended the concept of metamorphosis to animals, such as dogs and fish, that do not undergo this change. As in Study 1, more children endorsed the metamorphosis and species change trials for the fish (M = 0.87, SD = 1.06, out of 4), than the dog (M = 0.00, SD = 0.00). We fit a generalized linear mixed-effects model predicting the probability that children endorsed species change for the fish from test time (pretest versus posttest), diagram condition, question type, grade, use of general labels, the test time by diagram condition interaction, the grade by diagram condition interaction, and the general label by diagram condition interaction. We also included by-subject random intercepts and by-subject random slopes for the effect of test time and question type. Due to convergence issues, the random effects were not allowed to correlate.

We did not find an effect of test time or diagram condition, but we did find an interaction between the two. As can be seen in Figure 8, children who received the lesson with the bland diagram endorsed species changes for the fish item more at posttest (M = 0.27, SD = 0.45 out of 4) than at pretest (M = 0.13, SD = 0.34), and those who saw the lesson with the rich diagram endorsed these changes less at posttest (M = 0.17, SD = 0.37) than pretest (M = 0.22, SD = 0.42). This suggests that children who saw the bland diagram might overextend the concept of metamorphosis to species that do not undergo this change. Additionally, as in Study 1, children were more likely to endorse these changes for the lifespan questions (M = 0.23, SD = 0.42) than for the offspring questions (M = 0.17, SD = 0.37). We found no evidence for any other effects or interactions. See model summaries in Table 4.

Discussion

Study 2 shows that children in fourth and fifth grade benefitted from lessons with the bland diagram. Children in this study were more likely to recall the labels presented in the lesson if the lesson included the bland diagram. Additionally, there is some evidence that children transferred the information from the lesson better with the bland diagram. The bulk of the posterior distribution for the effect of diagram on transfer suggests that children transferred more with the bland diagram and that this effect might be more pronounced for children with low prior knowledge. The bland diagram might also have led to some inappropriate generalization, with children endorsing drastic changes for the fish, which does not undergo such changes.

General discussion

The studies presented in this paper highlight the importance of perceptual information for learning and generalization, and they also suggest important developmental changes in these effects. We found that children in first and second grade learned better from the lesson with the perceptually rich diagram. There was also some indication that these children generalized more from the lesson with the rich diagram. Conversely, children in fourth and fifth grade generalized (both correctly and incorrectly) more from the lesson with the bland diagram. The results for children in fourth and fifth grade are more similar to prior findings with adults with the identical lesson (Menendez et al., 2020). Thus, our studies suggest that the influence of perceptual richness on learning and generalization changes over the elementary school years.

The finding that bland representations did not lead to greater generalization for younger children is surprising. Studies in mathematics with similarly aged children show a consistent advantage of bland representations on transfer (Kaminski et al., 2008; Kaminski & Sloutsky, 2013). One possibility is that the rich diagram we used was not detrimental because the features were relevant (Rey, 2012; Siler & Willows, 2014), as all the details included in the rich diagram helped to identify the specific animal presented in the lesson. Additionally, it could be that the rich diagram was beneficial because the stimuli used at pretest were also rich, and therefore the richness could have served as a retrieval cue (Skulmowski & Rey, 2020a). However, none of these possibilities can explain why children in late elementary school generalized more with the bland diagram.

It is important to consider why children of different ages benefitted from different diagrams. As seen in Figure 6, performance with the rich diagram was very similar for the two groups of children. It was in performance with the bland diagram where we observed age-related changes. Thus, children in late elementary school learn better from bland representations than children in early elementary school—but both groups derive similar benefits from rich representations.

One possible explanation is that the number of bland representations used in educational materials increases over the elementary school years (Wiley et al., 2017; Menendez, Johnson, et al., 2020). As children receive more exposure to bland representations, they might develop skills for interpreting these representations. Theories of how people interpret visual representations argue that people have schemas that contain information about how the visualizations should look and what their elements represent (Padilla et al., 2018; Pinker, 1990). It is possible that due to the low frequency of bland representations in early elementary school, the young children did not have an appropriate schema for interpreting the bland diagram. The children in late elementary school might have had more experience with bland representations and might therefore have had a more appropriate schema to use. This could explain why performance with the rich representations does not change, but performance with the bland representation improves. Additionally, children's exposure to representations in general might also explain why prior research on mathematics learning has shown an advantage for bland representations, as bland representations might be more common in mathematics. Therefore, children might have appropriate schemas to interpret bland representations in mathematics but not in biology.

The idea that children have to learn how to interpret bland representations could also explain some of the benefits of instructional practices such as concreteness fading, in which children first see concrete representations and then are slowly introduced to blander or more abstract representations. The process of slowly fading representations might help children map between representations and understand which elements are important (Fyfe et al., 2014). Therefore, this fading procedure might be helping children create schemas of bland representations by using their schemas of rich representations as a scaffold, giving meaning and context to the bland representations. Children might make similar mappings as they are exposed to different types of visualizations at school. Future research should examine how manipulating the types of representations in children's environments influences how they learn with visual representations.

Our study also contributes to understanding of the development of biological reasoning. Prior work suggested that people rarely generalize the concept of metamorphosis to new or unfamiliar organisms (Herrmann et al., 2013). We found evidence supporting this infrequent generalization in our pretest data. At pretest, children rarely endorsed metamorphosis for ladybugs, an animal that was likely to be familiar to all of the children in our sample. This was the case even for fourth and fifth graders, who have likely had formal instruction on metamorphosis. However, we also found that children were open to generalizing this concept to other insects after a lesson. Furthermore, our lesson did not mention the appropriate scope of generalization, and many fourth and fifth grade children overextended this concept to animals that do not undergo this change, particularly when they had seen the bland diagram. Additionally, we saw that the extent to which children endorsed metamorphosis for ladybugs predicted whether they endorsed metamorphosis for other animals. This suggests that children used taxonomic categories to guide their generalization (i.e., if ladybugs go through metamorphosis, then other insects might also do so). Future studies should examine whether children generalize their knowledge to animals that are perceptually similar to insects, but do not belong to that category, such as spiders or centipedes. Additionally, future studies could also examine whether the semantic similarity of animals predicts how likely children are to generalize to those animals (Vales et al., 2020; Vales & Fisher, 2019).

It is important to acknowledge some limitations in our studies. First, children in our study likely had different experiences with formal lessons on metamorphosis. Although the Next

Generation Science Standards suggest that children should learn about metamorphosis by third grade, we do not know when this topic was covered in each child's curriculum. Therefore, some of the first and second grade students might have already had formal lessons, while some of the fourth and fifth grade students might have not had knowledge of metamorphosis before the lesson. We hoped to mitigate these differences in prior knowledge by controlling for pretest performance. Second, children completed the study on a one-on-one session in a research laboratory; therefore, children might have been highly motivated to pay attention to the lesson. Motivation might be lower in classroom settings. This could influence which type of visualization is more beneficial, as prior work in a laboratory setting suggests that rich visualizations lead to increased motivation, which in turn leads to better learning (Durik & Harackiewicz, 2007; Mayer et al., 2008; Sung & Mayer, 2012). Therefore, the influence of perceptual richness in a classroom setting might be different.

In spite of these limitations, our studies show that the perceptual richness of visual representations influences learning and transfer in different ways over development. By examining how children learn about a counterintuitive topic, metamorphosis, we were able to teach and assess children of different ages using the exact same materials—materials that have previously been used even with adults. This allowed us to see that children in early elementary school learn more with rich visual representations than with bland ones. The usefulness of bland visual representations increases over elementary school, such that fourth and fifth grade students generalize more from bland representations, which aligns with prior research with adults. This developmental trajectory mirrors the prevalence of bland representations in biology educational materials in elementary school, suggesting that children might benefit most from the types of visualizations they typically see in their environment. In sum, the effectiveness of visualizations in educational settings might depend, not only on the characteristics of the visualization, but also on changes over development. Thus, characteristics of the visualization interact with characteristics of the child to influence learning and transfer.

References

Alibali, M. W., Nathan, M. J., Wolfgram, M. S., Church, R. B., Jacobs, S. A., Johnson Martinez, C., & Knuth, E. J. (2014). How Teachers Link Ideas in Mathematics Instruction Using Speech and Gesture: A Corpus Analysis. *Cognition and Instruction*, *32*(1), 65–100. https://doi.org/10.1080/07370008.2013.858161

Angra, A., & Gardner, S. M. (2018). The Graph Rubric: Development of a Teaching, Learning, and Research Tool. *CBE—Life Sciences Education*, *17*(4), ar65. https://doi.org/10.1187/cbe.18-01-0007

Belenky, D. M., & Schalk, L. (2014). The Effects of Idealized and Grounded Materials on Learning, Transfer, and Interest: An Organizing Framework for Categorizing External Knowledge Representations. *Educational Psychology Review*, *26*(1), 27–50. https://doi.org/10.1007/s10648-014-9251-9

Butcher, K. R. (2006). Learning from text with diagrams: Promoting mental model development and inference generation. *Journal of Educational Psychology*, *98*(1), 182–197. https://doi.org/10.1037/0022-0663.98.1.182

Butler, F. M., Miller, S. P., Crehan, K., Babbitt, B., & Pierce, T. (2003). Fraction Instruction for Students with Mathematics Disabilities: Comparing Two Teaching Sequences. *Learning Disabilities Research & Practice*, *18*(2), 99–111. https://doi.org/10.1111/1540-5826.00066

Carbonneau, K. J., Wong, R. M., & Borysenko, N. (2020). The influence of perceptually rich manipulatives and collaboration on mathematic problem-solving and perseverance. *Contemporary Educational Psychology*, *61*, 101846. https://doi.org/10.1016/j.cedpsych.2020.101846

Castro-Alonso, J. C., Ayres, P., & Paas, F. (2016). Comparing apples and oranges? A critical look at research on learning from statics versus animations. *Computers & Education*, *102*, 234–243. https://doi.org/10.1016/j.compedu.2016.09.004

Cooper, J. L., Sidney, P. G., & Alibali, M. W. (2018). Who Benefits from Diagrams and Illustrations in Math Problems? Ability and Attitudes Matter: Diagrams and illustrations. *Applied Cognitive Psychology*, *32*(1), 24–38. https://doi.org/10.1002/acp.3371

De Bock, D., Deprez, J., Van Dooren, W., Roelens, M., & Verschaffel, L. (2011). Abstract or Concrete examples in Learning mathematics? A replication and elaboration of Kaminski, Sloutsky, and Heckler's study. *Journal for Research in Mathematics Education*, *42*(2), 109–126.

Durik, A. M., & Harackiewicz, J. M. (2007). Different strokes for different folks: How individual interest moderates the effects of situational factors on task interest. *Journal of Educational Psychology*, *99*(3), 597–610. https://doi.org/10.1037/0022-0663.99.3.597

Etta, R. A., & Kirkorian, H. L. (2019). Children's Learning From Interactive eBooks: Simple Irrelevant Features Are Not Necessarily Worse Than Relevant Ones. *Frontiers in Psychology*, *9*, 2733. https://doi.org/10.3389/fpsyg.2018.02733

Fisher, A. V., Godwin, K. E., & Seltman, H. (2014). Visual Environment, Attention Allocation, and Learning in Young Children: When Too Much of a Good Thing May Be Bad. *Psychological Science*, *25*(7), 1362–1370. https://doi.org/10.1177/0956797614533801

Flores, M. M. (2010). Using the Concrete-Representational-Abstract Sequence to Teach Subtraction With Regrouping to Students at Risk for Failure. *Remedial and Special Education*, *31*(3), 195–207. https://doi.org/10.1177/0741932508327467

Flynn, M. E., Guba, T. P., & Fyfe, E. R. (2020). ABBABB or 1212: Abstract language facilitates children's early patterning skills. *Journal of Experimental Child Psychology*, *193*, 104791. https://doi.org/10.1016/j.jecp.2019.104791

French, J. A., Menendez, D., Herrmann, P. A., Evans, E. M., & Rosengren, K. S. (2018). Cognitive constraints influence an understanding of life-cycle change. *Journal of Experimental Child Psychology*, *173*, 205–221. https://doi.org/10.1016/j.jecp.2018.03.018

Fyfe, E. R., McNeil, N. M., & Borjas, S. (2015). Benefits of "concreteness fading" for children's mathematics understanding. *Learning and Instruction*, *35*, 104–120. https://doi.org/10.1016/j.learninstruc.2014.10.004

Fyfe, E. R., McNeil, N. M., & Rittle-Johnson, B. (2015). Easy as ABCABC: Abstract Language Facilitates Performance on a Concrete Patterning Task. *Child Development*, *86*(3), 927–935. https://doi.org/10.1111/cdev.12331

Fyfe, E. R., McNeil, N. M., Son, J. Y., & Goldstone, R. L. (2014). Concreteness Fading in Mathematics and Science Instruction: A Systematic Review. *Educational Psychology Review*, *26*(1), 9–25. https://doi.org/10.1007/s10648-014-9249-3

Goldstone, R. L., & Sakamoto, Y. (2003). The transfer of abstract principles governing complex adaptive systems. *Cognitive Psychology*, *46*(4), 414–466. https://doi.org/10.1016/S0010-0285(02)00519-4

Goldstone, R. L., & Son, J. Y. (2005). The Transfer of Scientific Principles Using Concrete and Idealized Simulations. *Journal of the Learning Sciences*, *14*(1), 69–110. https://doi.org/10.1207/s15327809jls1401_4

Harp, S. F., & Mayer, R. E. (1998). How Seductive Details Do Their Damage: A Theory of Cognitive Interest in Science Learning. *Journal of Educational Psychology*, *90*(3), 414–434.

Hegarty, M., & Sims, V. K. (1994). Individual differences in mental animation during mechanical reasoning. *Memory & Cognition*, *22*(4), 411–430. https://doi.org/10.3758/BF03200867

Herrmann, P. A., French, J. A., DeHart, G. B., & Rosengren, K. S. (2013). Essentialist Reasoning and Knowledge Effects on Biological Reasoning in Young Children. *Merrill-Palmer Quarterly*, *59*, 198–220.

Kaminski, J. A., Sloutsky, V. M., & Heckler, A. F. (2008). The Advantage of Abstract Examples in Learning Math. *Science*, *320*(5875), 454–455. https://doi.org/10.1126/science.1154659

Kaminski, Jennifer A., & Sloutsky, V. M. (2013). Extraneous perceptual information interferes with children's acquisition of mathematical knowledge. *Journal of Educational Psychology*, *105*(2), 351–363. https://doi.org/10.1037/a0031040

Koedinger, K., Alibali, M., & Nathan, M. (2008). Trade-Offs Between Grounded and Abstract Representations: Evidence From Algebra Problem Solving. *Cognitive Science: A Multidisciplinary Journal*, *32*(2), 366–397. https://doi.org/10.1080/03640210701863933

Kruschke, J. K., & Liddell, T. M. (2018). The Bayesian New Statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic Bulletin & Review*, *25*(1), 178–206. https://doi.org/10.3758/s13423-016-1221-4

Lindner, M. A. (2020). Representational and decorative pictures in science and mathematics tests: Do they make a difference? *Learning and Instruction*, *68*, 101345. https://doi.org/10.1016/j.learninstruc.2020.101345

Lorenz, K. (1971). *Studies in Animal and Human Behaviour* (Vol. 2). Methuen.

Mayer, R. E. (2008). Applying the science of learning: Evidence-based principles for the design of multimedia instruction. *American Psychologist*, *63*(8), 760–769. https://doi.org/10.1037/0003-066X.63.8.760

Mayer, R. E. (2009). *Multimedia learning* (2nd ed). Cambridge University Press.

Mayer, R. E., Griffith, E., Jurkowitz, I. T. N., & Rothman, D. (2008). Increased interestingness of extraneous details in a multimedia science presentation leads to decreased learning. *Journal of Experimental Psychology: Applied*, *14*(4), 329–339. https://doi.org/10.1037/a0013835

McNeil, N. M., & Fyfe, E. R. (2012). "Concreteness fading" promotes transfer of mathematical knowledge. *Learning and Instruction*, *22*(6), 440–448. https://doi.org/10.1016/j.learninstruc.2012.05.001

Menendez, D., Mathiaparanam, O. N., Liu, D., Seitz, V., Alibali, M. W., & Rosengren, K. S. (2020). Representing Variability: The Case of Life Cycle Diagrams. *CBE—Life Sciences Education*, *19*(3), ar49. https://doi.org/10.1187/cbe.19-11-0251

Menendez, D., Rosengren, K. S., & Alibali, M. W. (2020). Do details bug you? Effects of perceptual richness in learning about biological change. *Applied Cognitive Psychology*, *34*(5), 1101–1117. https://doi.org/10.1002/acp.3698

Moreno, R., & Mayer, R. E. (1999). Cognitive Principles of Multimedia Learning: The Role of Modality and Contiguity. *Journal of Educational Psychology*, *91*(2), 358–368.

Padilla, L. M., Creem-Regehr, S. H., Hegarty, M., & Stefanucci, J. K. (2018). Decision making with visualizations: A cognitive framework across disciplines. *Cognitive Research: Principles and Implications*, *3*(1). https://doi.org/10.1186/s41235-018-0120-9

Pinker, S. (1990). A Theory of Graph Comprehension. In R. Freedle (Ed.), *Artificial intelligence and the future of testing* (pp. 73–126). Erlbaum.

Rau, M. A. (2017). Conditions for the Effectiveness of Multiple Visual Representations in Enhancing STEM Learning. *Educational Psychology Review*, *29*(4), 717–761. https://doi.org/10.1007/s10648-016-9365-3

Rey, G. D. (2012). A review of research and a meta-analysis of the seductive detail effect. *Educational Research Review*, *7*(3), 216–237. https://doi.org/10.1016/j.edurev.2012.05.003

Rosengren, K. S., Gelman, S. A., Kalish, C. W., & McCormick, M. (1991). As Time Goes By: Children's Early Understanding of Growth in Animals. *Child Development*, *62*, 1302-I320.

Sanchez, C. A., & Wiley, J. (2006). An examination of the seductive details effect in terms of working memory capacity. *Memory & Cognition*, *34*(2), 344–355. https://doi.org/10.3758/BF03193412

Schnotz, W., & Bannert, M. (2003). Construction and interference in learning from multiple representation. *Learning and Instruction*, *13*(2), 141–156. https://doi.org/10.1016/S0959-4752(02)00017-8

Schnotz, W., & Kürschner, C. (2008). External and internal representations in the acquisition and use of knowledge: Visualization effects on mental model construction. *Instructional Science*, *36*(3), 175–190. https://doi.org/10.1007/s11251-007-9029-2

Shepardson, D. P. (2002). Bugs, butterflies, and spiders: Children's understandings about insects. *International Journal of Science Education*, *24*(6), 627–643. https://doi.org/10.1080/09500690110074765

Siler, S. A., & Willows, K. J. (2014). Individual differences in the effect of relevant concreteness on learning and transfer of a mathematical concept. *Learning and Instruction*, *33*, 170–181. https://doi.org/10.1016/j.learninstruc.2014.05.001

Skulmowski, A., & Rey, G. D. (2018). Realistic details in visualizations require color cues to foster retention. *Computers & Education*, *122*, 23–31. https://doi.org/10.1016/j.compedu.2018.03.012

Skulmowski, A., & Rey, G. D. (2020a). Realism as a retrieval cue: Evidence for concreteness-specific effects of realistic, schematic, and verbal components of visualizations on learning and testing. *Human Behavior and Emerging Technologies*, 1–13.

Skulmowski, A., & Rey, G. D. (2020b). The realism paradox: Realism can act as a form of signaling despite being associated with cognitive load. *Human Behavior and Emerging Technologies*. https://doi.org/10.1002/hbe2.190

Son, J. Y., & Goldstone, R. L. (2009). Fostering general transfer with specific simulations. *Pragmatic & Cognition*, *17*(1), 1–42. https://doi.org/10.1075/p&c.17.1.01son

Sung, E., & Mayer, R. E. (2012). When graphics improve liking but not learning from online lessons. *Computers in Human Behavior*, *28*(5), 1618–1625. https://doi.org/10.1016/j.chb.2012.03.026

Trninic, D., Kapur, M., & Sinha, T. (2020). The Disappearing "Advantage of Abstract Examples in Learning Math." *Cognitive Science*, *44*(7). https://doi.org/10.1111/cogs.12851

Vales, C., & Fisher, A. V. (2019). When Stronger Knowledge Slows You Down: Semantic Relatedness Predicts Children's Co-Activation of Related Items in a Visual Search Paradigm. *Cognitive Science*, *43*(6). https://doi.org/10.1111/cogs.12746

Vales, C., States, S. L., & Fisher, A. V. (2020). Experience-Driven Semantic Differentiation: Effects of a Naturalistic Experience on Within- and Across-Domain Differentiation in Children. *Child Development*, *91*(3), 733–742. https://doi.org/10.1111/cdev.13369

Wiley, J. (2018). Picture this! Effects of photographs, diagrams, animations, and sketching on learning and beliefs about learning from a geoscience text. *Applied Cognitive Psychology*, 9–19. https://doi.org/10.1002/acp.3495

Wiley, J., Sarmento, D., Griffin, T. D., & Hinze, S. R. (2017). Biology Textbook Graphics and Their Impact on Expectations of Understanding. *Discourse Processes*, *54*(5–6), 463–478. https://doi.org/10.1080/0163853X.2017.1319655

Woodward, A. (1993). Do illustrations serve an instructional purpose in U.S. textbooks? In B. K. Britton, M. Binkley, A. Woodward, & M. R. Binkley (Eds.), *Learning from Textbooks: Theory and Practice* (pp. 115–134). Psychology Press.

Yzerbyt, V. Y., Muller, D., & Judd, C. M. (2004). Adjusting researchers' approach to adjustment: On the use of covariates when testing interactions. *Journal of Experimental Social Psychology*, *40*(3), 424–431. https://doi.org/10.1016/j.jesp.2003.10.001