Journal of Experimental Child Psychology 213 (2022) 105269

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

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

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

David Menendez [a,*], Karl S. Rosengren [b], Martha W. Alibali [a]

[a] Department of Psychology, University of Wisconsin–Madison, Madison, WI 53706, USA
[b] Department of Psychology and Department of Brain and Cognitive Sciences, University of Rochester, Rochester, NY 14627, USA

ARTICLE INFO

Article history:
Received 21 December 2020
Revised 19 July 2021
Available online 17 August 2021

Keywords:
Diagrams
Perceptual richness
Visualizations
Biological reasoning
Bayesian statistics
Biology education

ABSTRACT

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 regression analyses revealed that first and second graders learned more from the lesson with the perceptually rich diagram. In addition, fourth and fifth graders generalized more with the bland diagram, but these generalizations tended to be incorrect (i.e., generalizing metamorphosis to animals that do not undergo this type of change). These findings differ from prior research with adults, in which bland diagrams led to more correct generalizations, suggesting that the effect of perceptual richness on learning and generalization might change over development.

* Corresponding author.
E-mail address: dmenendez@wisc.edu (D. Menendez).

https://doi.org/10.1016/j.jecp.2021.105269

Introduction

Visualizations, such as graphs, diagrams, and pictures, are ubiquitous in educational materials. Visualizations are included in books (Menendez, Mathiaparanam, et al., 2020), presentations (Angra et al., 2018), tests (Lindner, 2020), and even classroom decorations (Fisher, Godwin, & Seltman, 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. Because visualizations can provide support for learning, they might be a particularly useful tool for teaching children about difficult or counterintuitive topics that might otherwise pose challenges for them to learn and generalize to new instances. In this research, we examined 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 on learning and generalization

Many studies have examined the influence of visual representations on learning and generalization. In general, these studies find that adding visual representations to a lesson leads to better learning and generalization (Mayer, 2008; Moreno & Mayer, 1999). However, not all visual representations are equally beneficial given that their effectiveness at promoting learning and generalization depends on characteristics of the representation (Rau & Matthews, 2017; Schnotz & Kürschner, 2008; Skulmowski & Rey, 2018).

One characteristic that has received a lot of attention in the psychological literature is the level of perceptual detail with which 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, & Alibali, 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, or depictive or as containing seductive, extraneous, or irrelevant details (Belenky & Schalk, 2014; Kaminski & Sloutsky, 2013; Menendez, Rosengren, & Alibali, 2020; Skulmowski & Rey, 2020). Likewise, line drawings have been described as abstract, idealized, generic, perceptually bland, symbolic, schematic, or sparse or as containing only relevant details (Butcher, 2006; Rey, 2012; Wiley, Sarmento, Griffin, & Hinze, 2017). Although the definitions 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 these related ideas are instantiated in research studies. For example, abstract representations (representations that depict general concepts rather than 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, Ayres, & Paas, 2016). In this research, we use the terms *perceptually rich* and *perceptually bland* because they represent our process in creating the visualizations. We define perceptual richness in terms of the number of visual features included in the representation. In addition, the terms *perceptually rich* and *perceptually bland* are useful when reviewing the literature because they can be applied to two-dimensional representations, such as photographs and diagrams, and also to three-dimensional representations, such as manipulatives (Carbonneau, Wong, & Borysenko, 2020).

Many studies have shown that perceptual richness leads to lower learning and transfer in children (Carbonneau, Wong, & Borysenko, 2020; Kaminski & Sloutsky, 2013) and adults (Butcher, 2006; Goldstone & Sakamoto, 2003; Menendez, Rosengren, & Alibali, 2020; Rey, 2012). For example, Kaminski & 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, & Alibali, 2020). Perceptually rich representations might be detrimental because they contain irrelevant details that learners need 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 as well, rich representations can be distracting, and this distraction can inhibit learning. In addition, rich representations can inhibit transfer because students may interpret them as overly specific. For example, after learning about metamorphosis with a rich life cycle diagram, people might infer that the lesson applies only to ladybugs, but if the lesson includes a bland diagram, people might infer that the lesson applies to other insects as well (Menendez, Rosengren et al., 2020).

However, some recent studies have shown that rich representations can promote learning, at least under some circumstances. Several studies have suggested that rich representations can promote learning if the details they contain are not distracting and instead are relevant to the task at hand (Belenky & Schalk, 2014; Siler & Willows, 2014; Trninic, Kapur, & Sinha, 2020). There is also support for the idea that rich representations are better for generalizing to other rich representations because the richness might serve as a retrieval cue (De Bock, Deprez, Dooren, Roelens, & Verschaffel, 2011; Skulmowski & Rey, 2020). In addition, children might learn and transfer better when lessons begin with rich representations and then slowly introduce bland representations. This procedure is referred to as *concreteness fading* (Fyfe, McNeil, Son, & Goldstone, 2014) or as the *concrete–representational–abstract* sequence (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) or the presence of other visualizations (Rau, 2017). One contextual factor explored in prior research is the generality of 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 accompanying language is general (Flynn, Guba, & Fyfe, 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).

In addition, 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, Sidney, & Alibali, 2018). Several studies have suggested that students with low prior knowledge benefit more from bland representations than from rich ones, whereas students with high prior knowledge perform similarly with both types of representations (Cooper, Sidney, & Alibali, 2018; Goldstone & Sakamoto, 2003). Taken together, this past work suggests that factors such as prior knowledge and the use of abstract language could moderate the effects of perceptual richness on learning and transfer.

Understanding of life cycle changes

The current studies focused on children's understanding of life cycle changes and in particular on the concept of metamorphosis. We focused on metamorphosis because prior research suggests that it is a difficult concept for people to grasp (Herrmann, French, DeHart, & Rosengren, 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 familiar organisms such as butterflies (French, Menendez, Herrmann, Evans, & Rosengren, 2018; Rosengren, Gelman, Kalish, & McCormick, 1991). Therefore, children and adults often do not think of drastic changes, such as metamorphosis, as a viable form of biological change, at least for most species (French, Menendez, Herrmann, Evans, & Rosengren, 2018; Rosengren, Gelman, Kalish, & McCormick, 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 (National Research Council, 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 such as 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). 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 simply get bigger with age (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. In addition, Menendez, Rosengren, et al., 2020 showed that adults could learn and generalize from a short lesson on metamorphosis that was designed for elementary school students. Therefore, focusing on metamorphosis enables us to examine the influence of diagrams on learning and transfer of knowledge across a variety of age ranges using identical materials.

Visual representations in biology education

Given that this research focused 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 textbooks about half of the visualizations were perceptually rich, and the proportion of perceptually rich visualizations decreased as the grade level of the textbooks 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 current studies, also suggest that there is variation in the perceptual richness of these diagrams. Mendendez, Mathiaparanam and colleagues (2020) analyzed life cycle diagrams found in textbooks, in 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 words.

The current studies

The current studies examined the effects of perceptual richness on children's learning and generalization of a counterintuitive biological concept—metamorphosis. We examined children's ability to generalize the concept of metamorphosis because prior work shows that people have difficulty in generalizing this concept beyond frogs and butterflies. Given that the Next Generation Science Standards suggest that children should learn about metamorphosis and other life cycle changes by third grade, Study 1 tested first- and second-grade students because they likely have had little exposure to 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 of French and colleagues (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 most people think does not undergo 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 (Rey, 2012). For this reason, it is difficult to know whether 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 to identify the animal displayed in the diagram as a ladybug. The bland diagram in our studies was created by removing details from the rich diagram. This makes the two diagrams more comparable, and more similar to each other, than in previous studies. Therefore, our studies provided a stringent test of the effects of adding or removing perceptual information because all the information was relevant.

At pretest, children were presented with a number of different animals and were asked about possible changes that could occur over the lifespan. We expected children to endorse change in size more than change in color, to endorse change in color more than metamorphosis, and to endorse metamorphosis more than change in species, and we expected that participants would endorse metamorphosis more for animals that actually undergo metamorphosis (French et al., 2018; Menendez, Rosengren, & Alibali, 2020). We expected that children would endorse metamorphosis for the ladybug more at posttest than at pretest because 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, & Alibali, 2020). This finding would show that children were able to learn from the lessons. However, children might learn better (i.e., endorse metamorphosis more for ladybugs) if they receive the lesson with the rich diagram, given that prior work shows that children learn well with rich materials (De Bock et al., 2011). Based on previous findings from Kaminski, Sloutsky and Heckler (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 more 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, Menendez, Herrmann, Evans, & Rosengren, 2018; Menendez, Rosengren, & Alibali, 2020). Finally, we explored whether children's prior knowledge and their use of general labels when recalling the animal in the lesson would moderate the effect of perceptual richness on transfer.

Study 1

Method

Participants

We recruited 76 children—38 first-grade students (M_{age} = 7.12 years, SD = 0.32) and 38 second-grade students (M_{age} = 8.09 years, SD = 0.29)—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). This sample size was selected to be comparable to other studies of the effect of visual representations on learning as well as other studies of children's biological reasoning (French, Menendez, Herrmann, Evans, & Rosengren, 2018; Herrmann, French, DeHart, & Rosengren, 2013; Kaminski & Sloutsky, 2013). The families had initially been recruited through local private and public schools, the local children's museum, and e-mails to employees at a large research university. The racial/ethnic makeup of the sample, as reported by 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 multiracial, and 5 who did not report race or ethnicity information. Families received $15 for their participation.

Design overview

The study was divided into three 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 min and focused on the life cycle of a ladybug. During the lesson, children saw either a perceptually rich or perceptually bland diagram. The posttest was similar to the pretest, except that it included more animals. The posttest contained three types of items: learning items (ladybugs and Asian beetles, which look similar to ladybugs), transfer items (non-ladybug insects, to which generalization is appropriate), and overextension items (non-insect animals, to which generalization is not appropriate).

Materials

All the stimuli, diagrams, and lesson scripts can be found on the Open Science Framework (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 et al., 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 5 animals at pretest (butterfly, ladybug, gray ladybug, fish, and dog) and 10 animals at posttest (ladybug, Asian beetle, firefly, stag beetle, ant, butterfly, praying mantis, fish, frog, and dog). Of these animals, only the fish and 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) was presented 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 accurately show the change. In color change trials, the target animal changed in both 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 questions, 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 because 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 forms for animals that do and do not go through metamorphosis for both question types can be seen in Fig. 1. A sample trial can be seen in Fig. 2.

The lesson focused on the life cycle of the ladybug and was delivered by a trained experimenter. The experimenter first presented the diagram and then gave the scripted, 2-min lesson. The diagram was either perceptually rich or perceptually bland, depending on the participant's condition assignment (see Fig. 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 could examine how far children generalize from the lesson.

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 lesson diagram, were presented on a desktop computer. The experimenter pointed at each form of the animal when asking each question.

Children first completed the pretest. Children then 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

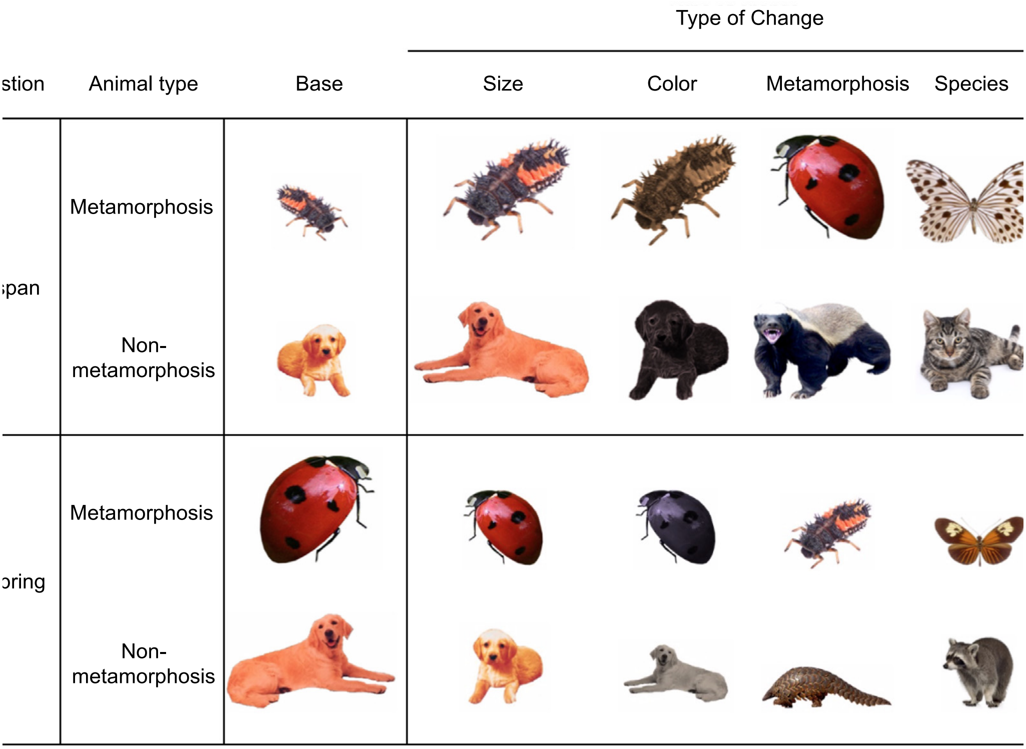

Fig. 1. Sample stimuli for both question types (lifespan and offspring) and animal types (metamorphosis and non-metamorphosis). The animals were always presented in pairs. The base was always presented on the left, and the target was always presented on the right.

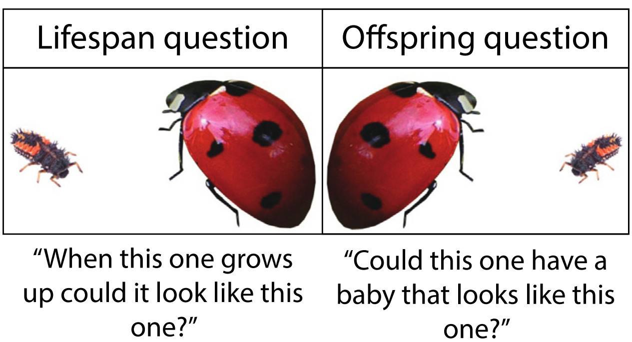

Fig. 2. Sample stimuli for both types of questions.

diagram. If children provided an incorrect label, the experimenter provided the correct label. After the recall questions, children completed the posttest.

While children completed the study, their parents filled out a demographic form on which they could report their children's age, gender, race/ethnicity, and grade in school.

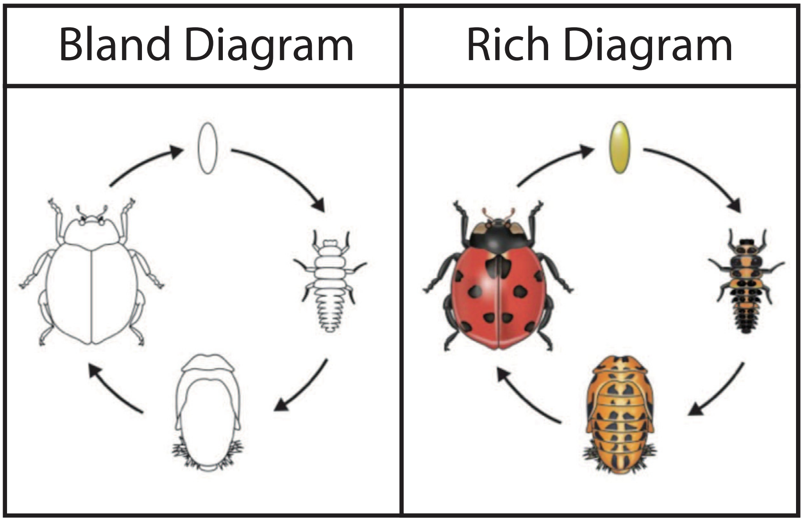

Fig. 3. Bland diagram (left panel) and rich diagram (right panel) used in the lesson. Everything else about the lesson was identical across conditions. Figures are available at https://osf.io/hfg38 under a CC-BY4.0 license (Menendez, 2019).

Coding

To examine children's responses to the recall task after the lesson, we transcribed their verbal responses and scored each label as correct or incorrect. Participants were coded as correct if they provided the label that was given in the lesson. Similar words, such as saying "pupis" instead of "pupa," were also counted as correct. Following research by Menendez, Rosengren et al. (2020), we coded participants' responses to the last stage as either *general* or *specific*. General labels were responses that did not mention the category of ladybug such as "adult," "beetle," "insect," and "adult stage." Specific labels were responses that mentioned the category of ladybug such as "ladybug" and "adult ladybug." All these labels were scored as correct, but they differ in whether children stated the information as specific to the ladybug or 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 children.

Results

Data analytic strategy

All the analyses presented in this article were done under a Bayesian framework using the *RStan* package (Stan Development Team, 2020) and *brms* package (Bürkner, 2017) in R (R Core Team, 2020) (for an overview of Bayesian data analysis, see Kruschke & Liddell, 2018). In all the models, the priors for the predictor variables were normal distributions with a mean of 0 and a standard deviation of 0.5. These priors have been called "skeptical priors" because they bias the models toward 0 (i.e., the predictor has no effect) and values close to it. For the random effects, we used the default priors in *brms*. That is, we used a half Student's *t* distribution with a mean of 0 and a standard deviation of 2.5 as our prior for the standard deviation for all the random effects. This is a weakly informative prior that allows for only positive numbers (because standard deviations can only be positive). For the correlation matrix of the random effects, we used *lkj*(1) as our prior, which places equal probability on all possible correlation matrices (this prior is uniform over the entire correlation matrix; individual correlation values are biased toward 0, with all values between −1 and 1 being possible). To conduct the analyses, we ran four Markov chains, with 5000 iterations each, with 1000 warm-up draws. To avoid overfitting, we performed leave-one-out cross-validation using the *loo* package (Vehtari, Gelman, & Gabry, 2017). During leave-one-out cross-validation, the model is trained on all the data except one observation, which is used to test the model's predictions. This process is repeated until every observation has been used to test the model. The average prediction error of the model is used to assess model fit.

The results of the leave-one-out cross-validation can be used to compare different models by comparing their expected log predictive density (*elpd*). Models with *elpd* differences less than 4 are considered to perform similarly in out-of-sample prediction. The model with the largest *elpd* is considered the best fitting model. For each analysis we present, we first fitted a model with predictors used in prior studies. This served as our baseline model. Then, we added diagram condition and interactions between diagram condition and pretest in subsequent models. We compare all these models using the *elpd*, and we present only the results of the best fitting model. If the best fitting model does not include diagram condition, it means that taking into account the diagram participants saw during the lesson does not lead to more accurate model predictions, suggesting that diagram had little effect on the outcome.

We take a similar approach to analyze pretest performance. We included type of change (size, color, metamorphosis, or species), animal type (metamorphosis or non-metamorphosis), and question type (lifespan or offspring) as predictors, but we did not allow them to interact in our baseline model. The subsequent models allowed for these predictors to interact in order to examine whether the interactions improved model fit.

All the models we fit throughout the article are logistic models with a Bernoulli link function. For each best fitting model, we report the odds ratio (*OR*), beta (the median of the posterior distribution in log odds), and 95% highest density intervals (HDI). The 95% HDI includes the most probable (also called credible) values for the effect of the predictor. These intervals are taken from the posterior distribution of the model so that every value that falls inside the interval is more likely than any value outside of it. The HDI does not need to have equal tails. Therefore, if zero is not included in the HDI, it means that zero is an unlikely value for the effect of the variable, suggesting that the predictor likely influences the outcome. If zero is included in the HDI, it suggests that zero is a likely value and therefore the predictor might not actually influence the outcome. In these cases, we can calculate what percentage of the posterior distribution is in the same direction as the beta. This is called the probability of direction, and it is useful to determine where zero falls in the distribution. If this probability is close to 50%, it suggests that zero is close to the center of the distribution (and that many likely values for the effect of the predictor are both positive and negative). If this probability is close to 95%, it suggests that although the effect might be zero, the bulk of the distribution suggests that the effect is in the same direction as the beta. To make reporting simpler, we report the probability of direction only when the value is higher than 85%.

First, we present an analysis of children's endorsement of life cycle changes at pretest as a partial replication of French and colleagues (2018). Then, we present the results of the recall task, both for whether participants correctly recalled the labels and whether they used general or specific labels for the last stage. Then, we present the results for learning, transfer, and overextension. The means reported are unadjusted mean proportions for each outcome. Model comparisons for all outcome variables can be found in Table 1. The analysis script can be found at https://osf.io/rqnem/?view_only=91450b4611044b3f95453db5ee6dc8f4.

Pretest performance

Our baseline model was a linear mixed effects model with a Bernoulli link function. We used whether participants answered "yes" or "no" on each trial as our outcome. We included grade (first or second), type of change (size, color, metamorphosis, or species), animal type (metamorphosis or non-metamorphosis), and question type (lifespan or offspring) as predictors, but we did not allow them to interact in the baseline model. We used dummy codes to examine the effect of type of change, and we set change in size as the reference category. We included by-participant random intercepts and by-participant random slopes for type of change, animal type, question type, and all interactions among the three. Subsequent models included interactions among type of change, animal type, and question type. As can be seen in Table 1, the best fitting model included the three-way interaction of type of change, animal type, and question type.

Table 1
Model comparisons for Study 1 and Study 2.

Model	Study 1		Study 2	
	Δ*elpd*	*SE*	Δ*elpd*	*SE*
Pretest				
Intercept + grade + change type + animal type + question type	−47.3	8.8	−22.6	5.5
Intercept + grade + change type × animal type + question type	−25.7	5.5	−12.6	3.1
Intercept + grade + change type + question type × animal type	−36.8	5.8	−24.2	5.5
Intercept + grade + change type × question type + animal type	−33.8	5.8	−20.3	5.1
Intercept + grade + change type × animal type + question type × animal type	−38.1	8.6	−12.4	2.8
Intercept + grade + change type × question type + question type × animal type	−34.4	7.9	−21.0	5.1
Intercept + grade + change type × question type + change type × animal type	−28.7	8.0	−8.9	2.3
Intercept + grade + change type × animal type + question type × animal type + change type × question type	−9.5	2.1	−9.0	1.9
Intercept + grade + change type × animal type × question type	0.0	–	0.0	–
Recall				
Intercept + grade + pretest score	0.0	–	−1.0	2.0
Intercept + grade + pretest score + diagram	−0.2	0.7	−0.2	1.2
Intercept + grade + pretest score × diagram	−0.8	0.7	0.0	–
Abstract label				
Intercept + grade + pretest score	−0.1	0.6	0.0	–
Intercept + grade + pretest score + diagram	0.0	–	−0.4	0.0
Intercept + grade + pretest score × diagram	−0.1	1.0	−1.2	0.2
Learning				
Intercept + grade + recall score + test time	−1.0	1.3	0.0	–
Intercept + grade + recall score + test time + diagram	−1.3	1.3	−0.3	0.6
Intercept + grade + recall score + test time × diagram	0.0	–	−0.5	0.5
Transfer				
Intercept + grade + abstract label + learning score + pretest score	0.0	–	0.0	–
Intercept + grade + abstract label + learning score + pretest score + diagram	−0.3	0.3	−0.1	0.6
Intercept + grade + abstract label + learning score + pretest score × diagram	−0.6	0.4	−0.1	0.6
Overextension				
Intercept + grade + test time	0.0	–	−0.6	0.8
Intercept + grade + test time + diagram	−0.1	0.7	−0.7	0.6
Intercept + grade + test time × diagram	−0.5	0.7	0.0	–

Note. The table reports the models fitted for each outcome measure. For each model, we report the change in expected log predictive density (Δ*elpd*) and standard error. A model with 0.0 as the Δ*elpd* means that this model was the best fitting model. Models with interactions also include all the relevant lower-order effects.

As hypothesized, children were more likely to endorse change in size (M = 0.68, SD = 0.47) than change in color (M = 0.35, SD = 0.48), OR = 0.26, b = −1.32 [−1.66, −0.98], more likely to endorse change in size than metamorphosis (M = 0.33, SD = 0.47), OR = 6.42, b = 1.86 [1.53, 2.19], more likely to endorse change in color than metamorphosis, OR = 0.67, b = −0.40 [−0.75, −0.06], and more likely to endorse metamorphosis than change in species (M = 0.06, SD = 0.24), OR = 0.11, b = −2.19 [−2.64, −1.79]. However, the pattern differed for metamorphosis and non-metamorphosis animals, as shown by interactions between animal type and the size change and metamorphosis contrast, OR = 7.61, b = 2.03 [1.53, 2.51], between animal type and the color change and metamorphosis contrast, OR = 0.20, b = −1.63 [−2.08, −1.19], and between animal type and the species change and metamorphosis contrast, OR = 0.54, b = −0.62 [−1.23, 0.01], with 97.3% of the posterior distribution being in the direction of b. To explore these interactions, we recentered our model at each type of change and looked at the simple effect of animal type. Children were more likely to endorse metamorphosis, OR = 4.35, b = 1.47 [1.15, 1.78], and change in species, OR = 1.54, b = 0.43 [0.05, 0.81], for animals that go through metamorphosis than for animals than do not go through metamorphosis. In addition, children were more likely to endorse change in species for lifespan questions for animals that undergo metamorphosis, OR = 2.53, b = 0.93 [0.41, 1.73]. See Fig. 4. There was no evidence for an effect of grade (first or second), OR = 0.90, b = −0.10 [−0.41, 0.21].

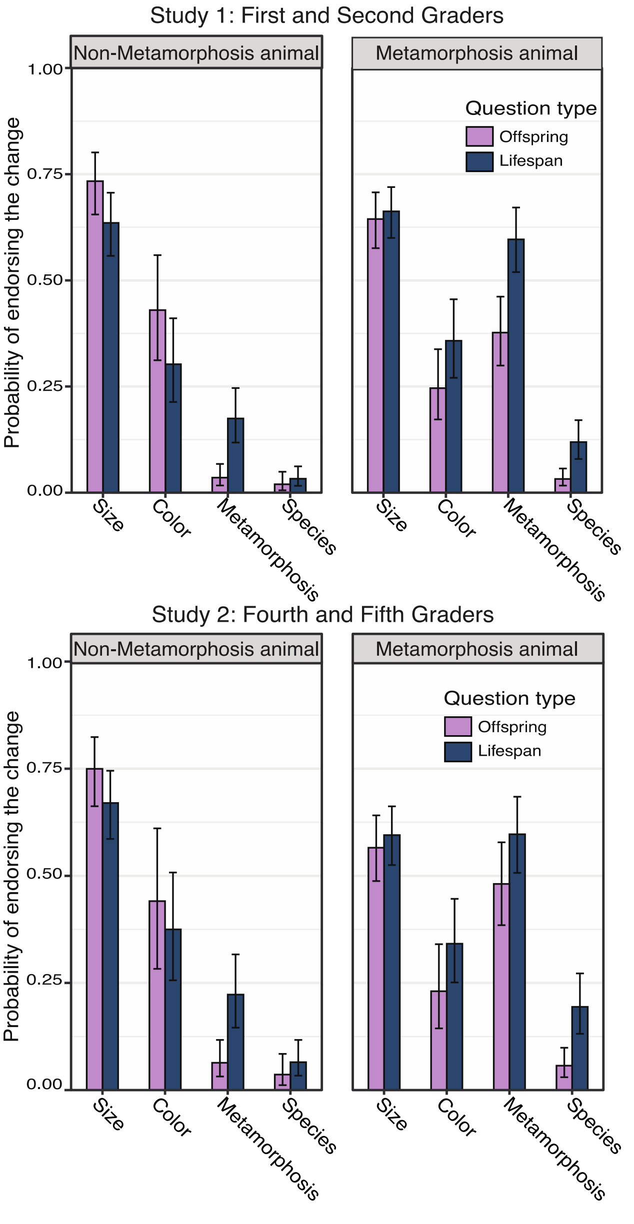

Fig. 4. Model predictions of the probability of endorsing each type of change, broken down by question type, for the best fitting model. The left panels show endorsements for animals that do not go through metamorphosis, and the right panels show endorsements for animals that go through metamorphosis. The top panels show the results for Study 1, and the bottom panels show the results for Study 2. The error bars represent the lower and upper bounds of the highest density interval. The model shows that children most frequently endorsed change in size. Even though children endorsed metamorphosis more for animals that undergo this type of change, they did so only about half the time (i.e., the probability of endorsement was near .50), suggesting that they do not consistently endorse metamorphosis even when it is appropriate.

Producing labels after the lesson

We fitted a linear mixed effects model that predicted whether children correctly recalled each label. We included pretest score and grade in our baseline model. We also included by-participant random intercepts. Subsequent models included diagram condition and the interaction between diagram condition and pretest score. As can be seen in Table 1, the best fitting model did not include diagram or the diagram by pretest interaction as predictors, suggesting that children in the rich condition (M = 3.35, SD = 0.79) and the bland condition (M = 3.26, SD = 0.86) correctly labeled similar numbers of stages. We found evidence for an effect of grade, such that second graders (M = 3.68, SD = 0.53) correctly labeled more stages than first graders (M = 2.95, SD = 0.90), OR = 2.66, b = 0.98 [0.42, 1.54]. We did not find evidence for an effect of pretest score, OR = 1.06, b = 0.06 [−0.11, 0.23].

We also examined whether the labels that children provided for the final stage were specific (e.g., "ladybug") or general (e.g., "insect," "adult"). We fitted a logistic regression predicting the probability of children providing a general label. In the baseline model, we included pretest score and grade as predictors. We also included diagram condition and the diagram condition by pretest score interaction. As can be seen in Table 1, the best fitting model included a main effect of diagram condition and suggests that children who saw the bland diagram were more likely to provide general labels than children who saw the rich diagram; however, the highest density interval included 0, suggesting that the evidence for this effect was weak, OR = 0.72, b = −0.33 [−1.03, 0.37]. We found evidence that children in second grade (M = 0.54, SD = 0.51) were more likely to provide a general label than children in first grade (M = 0.11, SD = 0.31), OR = 2.77, b = 1.02 [0.30, 1.74]. We found no evidence for an effect of pretest, OR = 1.06, b = 0.06 [−0.20, 0.31].

Learning

To examine whether children were more likely to endorse metamorphosis after the lesson, we compared children's responses to the ladybug items at pretest and posttest. We fitted a generalized linear mixed effects model with a bernoulli link function predicting children's endorsement of metamorphosis for the ladybug items. In the baseline model, we included test time (pretest or posttest), recall score (number of correct labels provided after the lesson), and grade. We also included by-participant random intercepts and by-participant random slopes for test time. In subsequent models, we included diagram condition and the diagram condition by test time interaction. As can be seen in Table 1, the best fitting model included the interaction between diagram condition and test time.

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), OR = 4.57, b = 1.52 [0.31, 0.92]. There was no main effect of diagram condition, OR = 0.95, b = −0.05 [−0.78, 0.67], but there was a test time by diagram condition interaction. As can be seen in Fig. 5, children who received the lesson with the rich diagram were more likely to endorse metamorphosis for ladybugs at posttest (M = 0.74, SD = 0.44) than children who received the lesson with the bland diagram (M = 0.58, SD = 0.50), OR = 2.69, b = 0.99 [0.25, 1.72]. We found some evidence for an effect of grade, OR = 1.70, b = 0.53 [−0.19, 1.25], with 92.42% of the posterior distribution suggesting that second graders (M = 0.56, SD = 0.50) were more likely to endorse metamorphosis for ladybugs after the lesson than first graders (M = 0.39, SD = 0.49). We also found some evidence for an effect of the number of labels correctly recalled, OR = 1.46, b = 0.38 [−0.18, 0.94], with 91.13% of the posterior distribution suggesting that children who recalled more labels after the lesson were more likely to endorse metamorphosis for ladybugs than those who recalled fewer labels.

Transfer

To examine children's generalization, we fitted a generalized linear mixed effects model predicting children's endorsement of metamorphosis for non-ladybug insects. In the baseline model, we included pretest and grade as predictors. 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). In addition, 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. We also included by-participant random intercepts. In subsequent models, we included diagram condition and the diagram condition by pretest interaction.

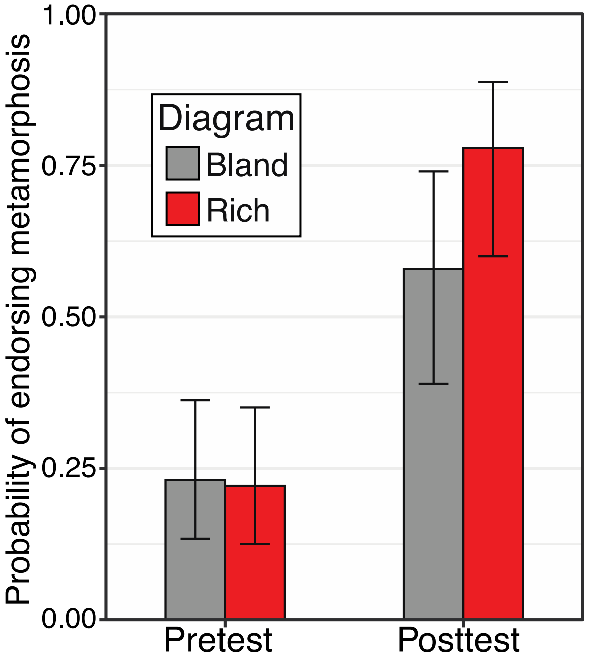

Fig. 5. Model predictions of the probability of endorsing metamorphosis for ladybugs at pretest (left set of bars) and posttest (right set of bars), for children who received the lesson with the bland life cycle diagram (gray [left] bars) and rich life cycle diagram (red [right] bars), for the best fitting model. The error bars represent the lower and upper bounds of the highest density interval. Higher values on the *y* axis indicate a higher probability of correctly endorsing metamorphosis for ladybugs. The model suggests that for Study 1 overall, children learned from the lesson, but those who saw the rich diagram learned more.

As can be seen in Table 1, the best fitting model did not include a main effect of diagram or the diagram by pretest interaction. This suggests that, contrary to our prediction, there was no evidence that children who saw the bland diagram (M = 0.56, SD = 0.50) were more likely to transfer than children who saw the rich diagram (M = 0.62, SD = 0.49). We did find evidence that as children's pretest scores increased, children were more likely to transfer, OR = 1.36, b = 0.31 [0.14, 0.49]. In addition, children who endorsed metamorphosis more for the learning items were more likely to endorse metamorphosis for the transfer items, OR = 1.38, b = 0.32 [0.10, 0.54]. We did not find evidence for an effect of grade, OR = 1.25, b = 0.22 [−0.33, 0.78], or for an effect of general labels, OR = 1.32, b = 0.28 [−0.29, 0.85].

Overextension

We also examined 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, both the metamorphosis and species change trials are nonbiological 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 fitted a generalized linear mixed effects model with a Bernoulli link function predicting the probability that children endorsed metamorphosis for the fish. In the baseline model, we included test time (pretest or posttest) and grade as predictors. We also included by-participant random intercepts and by-participant random slopes for the effect of test time and allowed them to correlate. In subsequent models, we included diagram condition and the interaction between diagram condition and test time.

As can be seen in Table 1, the best fitting model did not include a main effect of diagram or the diagram by test time interaction. We also did not find evidence for an effect of test time, $OR = 1.21$, $b = 0.19$ [−0.32, 0.66], or an effect of grade, $OR = 1.06$, $b = 0.06$ [−0.47, 0.59].

Discussion

We examined whether the perceptual richness of diagrams influenced first and second graders' 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, which is contrary to findings of previous work with adults (Menendez, Rosengren et al., 2020). This suggests that the effects of perceptual richness on children's learning and generalization are different from those on adults' learning and generalization.

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 students. We used the same lessons and testing materials as in Study 1 and in previous research with adults (Menendez, Rosengren 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, all the students should have had relatively recent exposure to the concept of metamorphosis. In addition, during these later school years, educational materials include more bland representations (Menendez, Johnson, et al., 2020). Therefore, we expected that fourth and fifth graders might benefit from the bland diagram. All other predictions were the same as in Study 1.

Study 2

Method

Participants

We recruited 53 children—30 fourth-grade students (M_{age} = 10.38 years, SD = 0.50) and 23 fifth-grade students (M_{age} = 10.88 years, SD = 0.60)—from the same database used in Study 1 (27 boys and 26 girls). We initially intended to collect the same number of participants as in Study 1, but we had to stop data collection due to the onset of the COVID-19 pandemic. The racial/ethnic makeup 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 multiracial, and 1 who reported another racial/ethnic category. Families received $15 for participating in the study.

Materials and procedure

The design, materials, and procedure were identical to those in 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 article, we do not discuss responses to these questions here. All data, materials, and analysis scripts can be found at https://osf.io/rqnem/?view_only=91450b4611044b3f95453db5ee6dc8f4.

Results

Data analysis

We used the same data analytic approach and fitted the same models as in Study 1. First, we present the results for children's endorsement of life cycle changes before the lesson. Then, we present the results for the recall task, both for whether participants correctly recalled the labels and for whether they used general or specific labels for the last stage. Then, we present the results for learning, transfer, and overextension. The means reported are unadjusted mean proportions for each outcome. Model comparisons for all outcome variables can be found in Table 1.

Pretest performance

As can be seen in Table 1, as in Study 1, the best fitting model of pretest performance included the three-way interaction of type of change, animal type, and question type. As in Study 1, children were more likely to endorse change in size (M = 0.64, SD = 0.48) than change in color (M = 0.35, SD = 0.48), OR = 0.33, b = −1.12 [−1.52, −0.72], and more likely to endorse change in size than metamorphosis (M = 0.38, SD = 0.49), OR = 0.29, b = −1.22 [−1.58, −0.85]. However, in this study, there was no difference in endorsement of change in color and metamorphosis, OR = 0.87, b = −0.14 [−0.58, 0.28]. Children were also more likely to endorse metamorphosis than change in species (M = 0.11, SD = 0.31), OR = 0.15, b = −1.90 [−2.34, −1.50]. 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, OR = 9.68, b = 2.27 [1.76, 2.78], and between animal type and the color change and metamorphosis contrast, OR = 0.21, b = −1.57 [−2.08, −1.02]. To explore these interactions, we recentered our model for each type of change and looked at the simple effect of animal type. Children were more likely to endorse metamorphosis, OR = 3.67, b = 1.30 [0.96, 1.64], and change in species, OR = 1.52, b = 0.42 [0.03, 0.81], and were less likely to endorse change in size, OR = 0.66, b = −0.41 [−0.72, −0.09], for animals that go through metamorphosis than for animals that do not go through metamorphosis. In addition, as in Study 1, children were more likely to endorse change in species for the lifespan questions for animals that undergo metamorphosis, OR = 2.77, b = 1.02 [0.20, 1.83]. See Fig. 4. There was no evidence for an effect of grade, OR = 0.99, b = −0.01 [−0.35, 0.33].

Producing labels after the lesson

As can be seen in Table 1, unlike Study 1, the best fitting model included diagram and the diagram by pretest score interaction, but we did not find evidence for a main effect of diagram, OR = 1.04, b = 0.04 [−0.89, 0.97], or an effect of pretest score, OR = 1.04, b = 0.04 [−0.27, 0.37]. We did find some evidence for a diagram by pretest score interaction, OR = 0.84, b = −0.17 [−0.40, 0.04], with 94.10% of the posterior distribution being in the same direction as the beta. Children with high prior knowledge were more likely to correctly recall the labels if they saw the bland diagram than if they saw the rich diagram. However, children with low prior knowledge were not affected by the diagram condition. See Fig. 6. We found no evidence for an effect of grade, OR = 1.39, b = 0.33 [−0.41, 1.07].

We also sought to predict whether children used general labels to describe the final stage. As can be seen in Table 1, the best fitting model did not include an effect of diagram or the diagram by pretest interaction. We also did not find an effect of grade, OR = 1.09, b = 0.09 [−0.68, 0.86]. We found that as pretest scores increased, children were more likely to provide a general label, OR = 1.52, b = 0.42 [0.08, 0.79].

Learning

We also examined whether children were more likely to endorse metamorphosis for the ladybug items after the lesson. As can be seen in Table 1, unlike Study 1, the best fitting model did not include a main effect of diagram or a diagram by test time interaction, suggesting that the diagram condition did not influence learning. 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), OR = 6.55, b = 1.88 [1.05, 2.58]. We did not find evidence for an effect of number of labels recalled, OR = 1.38, b = 0.32 [−0.28, 0.91], or an effect of grade, OR = 1.05, b = 0.05 [−0.60, 0.71].

Transfer

We next examined children's endorsement of metamorphosis for the non-ladybug insect items. As in Study 1, the best fitting model did not include an effect of diagram or the diagram by pretest score interaction, suggesting that children generalized similarly with the bland diagram (M = 0.80, SD = 0.40) and with the rich diagram (M = 0.76, SD = 0.43). See Table 1. There was no indication of an effect of pretest score, OR = 1.02, b = 0.02 [−0.16, 0.20], grade, OR = 0.75, b = −0.29 [−0.80, 0.24], or the use of general labels, OR = 1.18, b = 0.17 [−0.38, 0.71]. There was some indication of

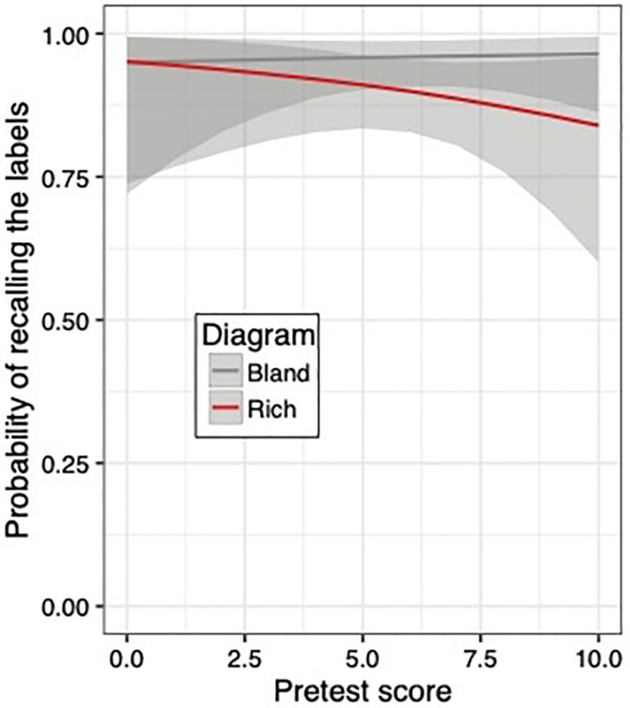

Fig. 6. Model predictions of the probability of correctly recalling a label after the lesson by pretest scores (on the *x* axis) for children who received the lesson with the bland life cycle diagram (gray [upper] line) and those who received the lesson with the rich life cycle diagram (red [lower] line), for the best fitting model. The error bands represent the lower and upper bounds of the highest density interval. Higher values on the *y* axis indicate a higher probability of recalling the labels. The model suggests that for Study 2, as prior knowledge (pretest score) increased, children who saw the bland diagram were increasingly more likely to recall the correct labels than children who saw the rich diagram.

an effect of learning score, *OR* = 1.27, *b* = 0.24 [−0.12, 0.60], with 90.67% of the posterior distribution being in the same direction as the beta, suggesting that children who endorsed metamorphosis more for ladybugs might also be more likely to endorse metamorphosis for non-ladybug insects.

Overextension

As in Study 1, we also examined whether children overextended the concept of metamorphosis to animals that do not undergo this change, such as dogs and fish. Also as in Study 1, more children endorsed the metamorphosis and species change trials for the fish (*M* = 0.87 out of 4, *SD* = 1.06) than for the dog (*M* = 0.00 out of 4, *SD* = 0.00). As can be seen in Table 1, in the best fitting model, we did not find evidence for an effect of test time, *OR* = 1.08, *b* = 0.08 [−0.58, 0.71], or diagram condition, *OR* = 1.26, *b* = 0.23 [−0.33, 0.80]. However, there was some evidence for an interaction between test time and diagram, *OR* = 0.63, *b* = −0.46 [−1.18, 0.26], with 89.53% of the posterior distribution being in the same direction as the beta. As can be seen in Fig. 7, children who received the lesson with the bland diagram endorsed species changes for the fish item more at posttest (*M* = 0.27 out of 4, *SD* = 0.45) than at pretest (*M* = 0.13 out of 4, *SD* = 0.34), and those who saw the lesson with the rich diagram endorsed these changes less at posttest (*M* = 0.17 out of 4, *SD* = 0.37) than at pretest (*M* = 0.22 out of 4, *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.

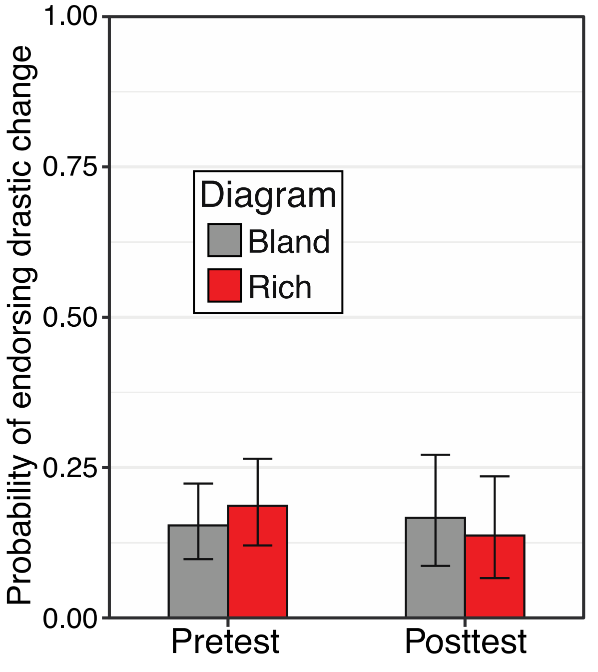

Fig. 7. Model predictions of the probability of endorsing drastic life cycle changes for the fish items at pretest (left set of bars) and at posttest (right set of bars), for children who received the lesson with the bland life cycle diagram (gray [left] bars) and the rich life cycle diagram (red [right] bars), for the best fitting model. The error bars represent the lower and upper bounds of the highest density interval. Higher values on the *y* axis indicate a higher probability of incorrectly endorsing metamorphosis for fish. The model suggests that for Study 2, children who saw the rich diagram endorsed metamorphosis for fish less at posttest than at pretest, whereas those who saw the bland diagram endorsed metamorphosis for fish more at posttest than at pretest.

Discussion

Study 2 shows that fourth and fifth graders benefitted from lessons with the bland diagram. Children with high prior knowledge in this study were more likely to recall the labels presented in the lesson if the lesson included the bland diagram. 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 article suggest that there may be developmental changes in the importance of perceptual information for learning and generalization. In Study 1, we found that first- and second-grade students learned better from the lesson with the perceptually rich diagram. In Study 2, fourth- and fifth-grade students were more likely to recall labels (for those with high prior knowledge) and more likely to incorrectly generalize from the lesson with the bland diagram. These results are different from those of previous studies with adults with the identical lesson (Menendez, Rosengren 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 correct generalization for children is surprising. Studies in mathematics with similarly aged children show a consistent advantage of bland representations on transfer (Kaminski et al., 2008). One possibility is that the rich diagram we used was not detrimental because the features were relevant (Rey, 2012; Siler & Willows, 2014), given that all the details included in the rich diagram helped to identify the specific animal presented in the lesson. However, none of these possibilities can explain why children overgeneralized more with the bland diagram.

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, Creem-Regehr, Hegarty, & Stefanucci, 2018). It is possible that due to the low frequency of bland representations that children encounter during their early elementary school years, the first and second graders did not have an appropriate schema for interpreting the bland diagram, and thus it did not improve their generalization. In addition, children's exposure to representations in general might also explain why prior research on mathematics learning has shown an advantage for bland representations, given that 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 need 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 aspects of the representations might help children to map between the representations and understand which elements are important (Fyfe, McNeil, Son, & Goldstone, 2014). Therefore, this fading procedure might be helping children to create schemas for 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 the children in our sample. This was the case even for fourth and fifth graders, who presumably had 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 graders overextended this concept to an animal that does not undergo this change (the fish), particularly if they had seen the bland diagram. In addition, 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 and centipedes. In addition, future studies could also examine whether the semantic similarity of animals predicts how likely children are to generalize to those animals (Vales & Fisher, 2019).

It is important to acknowledge some limitations of these studies. First, children may have 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 graders might have already had formal lessons, whereas some of the fourth and fifth graders might not have had knowledge of metamorphosis before participating in our study. We hoped to mitigate these differences in prior knowledge by controlling for pretest performance. Second, because we conducted the two studies separately, we cannot determine whether the differences are due to age rather than some other factor. We attempted to make the studies as comparable as possible by having the same experimenter conduct both studies, but we still cannot ascertain whether age is the critical factor that explains the observed differences in performance. Third, the influence of perceptual richness in a classroom setting might be different from what we found in the current studies. Children completed these studies in one-on-one sessions in a research laboratory; therefore, they 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, given that prior work in a laboratory setting has suggested that rich visualizations lead to increased motivation, which in turn leads to better learning (Durik & Harackiewicz, 2007; Sung & Mayer, 2012). Finally, although our results suggest that the type of visualization influenced some of the results, it is worth highlighting that these effects were small, given that the predictive power of the models was not greatly affected when these variables were removed. Thus, although we show some effects of perceptual richness, the effects for children might be smaller than those previously reported for adults.

In spite of these limitations, our studies show that the perceptual richness of visual representations influences learning and generalization 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 first and second graders learned more with a rich visual representation than with a bland one. Fourth and fifth graders overgeneralized more with a bland visual representation than with a rich one. This is different from previous findings with adults, who correctly generalized more with a bland visual representation than with a rich one. This developmental trajectory mirrors the prevalence of bland representations in biology educational materials in elementary school, potentially suggesting that children might benefit most from the types of visualizations they typically see in their everyday environments. In sum, the effectiveness of visualizations in educational settings might depend both on the characteristics of the visualizations and also on changes that occur over development.

Acknowledgments

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 the 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). In addition, this research was partially supported by a grant from the National Science Foundation (MSN21179). Finally, this research was also supported by an undergraduate Senior Thesis Award to the first author from the Department of Psychology at the University of Wisconsin–Madison. We thank Kathryn Hartfield, Erin Condon, Olympia Mathiaparanam, and Vienne Seitz for their help with data collection.

References

Angra, A., Gardner, S. M., & Knight, J. (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.

Bürkner, P. C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software, 80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01.

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.

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. *Applied Cognitive Psychology, 32*(1), 24–38. https://doi.org/10.1002/acp.3371.

De Bock, D., Deprez, J., Dooren, W. V., 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.

Evans, E. M. (2001). Cognitive and contextual factors in the emergence of diverse belief systems: Creation versus evolution. *Cognitive Psychology, 42*(3), 217–266. https://doi.org/10.1006/cogp.2001.0749.

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., & Rittle-Johnson, B. (2015). Easy as ABCABC: Abstract language facilitates performance on a concrete patterning task. *Child Development, 86*, 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.

Hegarty, M., & Sims, V. K. (1994). Individual differences in mental animation during mechanical reasoning. *Memory & Cognition, 22*(4), 411–430.

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*(2), 198–220.

Kaminski, J. 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.

Kaminski, J. A., Sloutsky, V. M., & Heckler, A. F. (2008). The advantage of abstract examples in learning math. *Science, 320*, 454–455.

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.

Mayer, R. E. (2008). Applying the science of learning: Evidence-based principles for the design of multimedia instruction. *American Psychologist, 63*(8), 760–769.

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.

Menendez, D. (2019). Perceptually rich and bland life-cycle diagrams. Retrieved from https://osf.io/hfg38.

Menendez, D., Johnson, T. M., Hassett, R., Haut, A., Mathiaparanam, O. N., Alibali, M. W., & Rosengren, K. S. (2020). Characteristics of visualizations and texts in elementary school biology books. In Poster presented at the 42nd annual conference of the Cognitive Science Society, virtual meeting.

Menendez, D., Mathiaparanam, O. N., Liu, D., Seitz, V., Alibali, M. W., Rosengren, K. S., & Tanner, K. (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. https://doi.org/10.1037/0022-0663.91.2.358.

National Research Council (2013). *Next generation science standards: For states, by states*. Washington, DC: National Academies Press.

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), 29. https://doi.org/10.1186/s41235-018-0120-9.

R Core Team (2020). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

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.

Rau, M. A., & Matthews, P. G. (2017). How to make "more" better? Principles for effective use of multiple representations to enhance students' learning about fractions. *ZDM, 49*(4), 531–544. https://doi.org/10.1007/s11858-017-0846-8.

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*(6), 1302. https://doi.org/10.2307/1130808.

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.

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.

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. (2020). The realism paradox: Realism can act as a form of signaling despite being associated with cognitive load. *Human Behavior and Emerging Technologies, 2*(3), 251–258. https://doi.org/10.1002/hbe2.190.
Son, J. Y., & Goldstone, R. L. (2009). Fostering general transfer with specific simulations. *Pragmatics & Cognition, 17*(1), 1–42. https://doi.org/10.1075/p&c.17.1.01son.
Stan Development Team. (2020). RStan: The R interface to Stan. R package Version 2.21.1. URL: http://mc-stan.org.
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.v44.710.1111/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.2019.43.issue-610.1111/cogs.12746.
Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing, 27*(5), 1413–1432. https://doi.org/10.1007/s11222-016-9696-4.