Optional ERIC Coversheet — Only for Use with U.S. Department of Education Grantee Submissions

This coversheet should be completed by grantees and added to the PDF of your submission if the information required in this form **is not included on the PDF to be submitted**.

INSTRUCTIONS

- Before beginning submission process, download this PDF coversheet if you will need to provide information not on the PDF.
- Fill in all fields—information in this form **must match** the information on the submitted PDF and add missing information.
- Attach completed coversheet to the PDF you will upload to ERIC [use Adobe Acrobat or other program to combine PDF files]—do not upload the coversheet as a separate document.
- Begin completing submission form at https://eric.ed.gov/submit/ and upload the full-text PDF with attached coversheet when indicated. Your full-text PDF will display in ERIC after the 12-month embargo period.

GRANTEE SUBMISSION REQUIRED FIELDS

Title of article, paper, or other content

Cues to generality: Integrating linguistic and visual information when generalizing biological information

All author name(s) and affiliations on PDF. If more than 6 names, ERIC will complete the list from the submitted PDF.

Last Name, First Name	Academic/Organizational Affiliation	ORCID ID
Menendez, David	University of California-Santa Cruz	

Publication/Completion Date—(if *In Press,* enter year accepted or completed) 2023

Check type of content being submitted and complete one of the following in the box below:

- ☑ If article: Name of journal, volume, and issue number if available
- ☐ If paper: Name of conference, date of conference, and place of conference
- ☐ If book chapter: Title of book, page range, publisher name and location
- ☐ If book: Publisher name and location
- ☐ If dissertation: Name of institution, type of degree, and department granting degree

Journal of Educational Psychology, 115, 8

DOI or URL to published work (if available) 10.1037/edu0000812

Acknowledgement of Funding— Grantees should check with their grant officer for the preferred wording to acknowledge funding. If the grant officer does not have a preference, grantees can use this suggested wording (adjust wording if multiple grants are to be acknowledged). Fill in Department of Education funding office, grant number, and name of grant recipient institution or organization.

"This work was supported by U.S. Department of Education **[Office name]** Institute of Education Scie through **[Grant number]** R305B1500 to **Institution]** University of W. The opinions expressed are those of the authors and do not represent views of the **[Office name]** Institute of Education Scie or the U.S. Department of Education.

Cues to generality:

Integrating linguistic and visual information when generalizing biological information

David Menendez

University of California-Santa Cruz

Acknowledgements

I would like to thank Taylor Johnson, Ashley Haut, Ryan Hassett, and Olympia Mathiaparanam for their help collecting and coding the books in Study 1. I would also like to thank Karl Rosengren, Martha Alibali, Percival Matthews, Karen Schloss and Kristin Shutts for their feedback on the design and comments on earlier versions of this manuscript and Susan Gelman for feedback on earlier drafts. 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. Studies 1-5 were not pre-registered. The design, hypotheses and analyses plans for Studies 6-8 were pre-registered; Study 6 see https://osf.io/cm3w9/?view_only=c8e367e945b54fa9bde937fe6920976a; Study 7 see https://osf.io/rbguz?view_only=c394c55989d84774a47ef1b21be6c0a3; Study 8 see https://osf.io/5trbc/?view_only=4af865c07bfb48c7bb46f4bbedbbcd8e.

Abstract

During instruction, students are typically presented with new information through several modalities, such as through language and images. Students need to attend to these different modalities and integrate the information in both in order to learn and generalize from instruction. Many studies have shown that the features of each modality, such as the use of generic noun phrases or perceptually bland visualizations, influence how much students generalize. However, few studies have manipulated both the linguistic and visual information to examine how students integrate the two modalities and how they generalize when the modalities cue to different levels of generalization. Study 1 examines what combinations of linguistic and visual information are common in elementary school science books. Studies 2-6 show that undergraduate students rely primarily on the linguistic information when generalizing. Study 7 reduced the possible split of visual attention by reading out the text for participants and shows that undergraduate students generalize more broadly when the information in either modality promotes generalization, but their effect does not compound. Study 8 shows that elementary school students generalize more broadly when the linguistic information is broad, but the visual information is rich. These results suggest that students across ages use linguistic features similarly to guide their generalizations, but how they integrate the linguistic and visual information changes with age. Based on these findings, I propose the cues to generality hypothesis, as an account of how students use information in lessons to determine how far to generalize.

Keywords: Multimedia learning, Generic language, Perceptual richness, Generalization, Text-picture integration.

Educational impact and implications statement

Lessons often combine pictures and text, and the characteristics of both of these elements influence generalization. Undergraduate students generalize more broadly when the text contains generic noun phrases or the picture is bland, but elementary school students generalize more broadly when the text contains generic noun phrases, and the image is rich. This shows that how students use images to support their generalizations changes with age.

Cues to generality:

Integrating linguistic and visual information when generalizing biological information

During instruction, students are typically presented with new information through several modalities, such as through text and images (Woodward, 1993). In order to effectively learn and generalize from a lesson students need to attend to these different modalities and combine the information that is conveyed in each modality (Mayer, 2009; Schnotz, 2014). Once they have integrated this information, students have to decide how far to generalize to other exemplars that were not included in the lesson. Although many studies have investigated how characteristics of the visualization (Butcher, 2006; Cooper et al., 2018; Goldstone & Sakamoto, 2003; Kaminski et al., 2008; Skulmowski, 2022; Son & Goldstone, 2009) and the language (Brandone & Gelman, 2009; Flynn et al., 2020; Fyfe et al., 2015; Hoicka et al., 2021; Leshin et al., 2021) influence how students generalize, few studies have examined how these characteristics may interact. In this paper, I focus on one linguistic feature, generic noun phrases, and one visual feature, perceptual richness, that have been shown to independently promote generalization, to examine how students combine linguistic and visual information to constrain or broaden their generalizations.

Generic language

There are several linguistic features that influence generalization, such as whether the text includes irrelevant information (Garner, 1992; Garner et al., 1989, 1991), whether the language is general or specific (Fyfe et al., 2015; Son & Goldstone, 2009), and whether it includes a generic noun phrase (Hollander et al., 2002). In this paper, I focus on the use of generic (e.g., “*Birds* lay eggs”) or non-generic (e.g., “*This bird* lays eggs”) noun phrases. Generic noun phrases are produced and comprehended from an early age (Brandone et al., 2012; Brandone & Gelman, 2009), and common in children’s and adults’ environments (Gelman et al., 1998, 2013) particularly in pedagogical contexts (Gelman et al., 2013). Additionally, previous work has shown that generic noun phrases (referred to as "generics" hereafter) promote generalization to all the members of a category (Brandone et al., 2012; Brandone & Gelman, 2009; Cimpian & Markman, 2009; Kochari et al., 2020; Leslie, 2007; Noyes & Keil, 2019; Tasimi et al., 2017), even when this generalization is not appropriate (e.g., only female birds lay eggs). Thus, generic statements are common in educational settings and tend to promote generalization regardless of whether there is a visualization present.

Perceptual richness

One particular feature of visualizations that has received a lot of attention in the psychological literature is their perceptual richness. Perceptual richness refers to the number of visual features included in a visualization. Perceptual richness is related to other characteristics of visualizations, such as concreteness, that have been found to influence generalization (see Castro-Alonso et al., 2016; and Menendez et al., 2022 for further discussion). Here, I use the term perceptual richness as it aligns more directly with how the stimuli was constructed. Many studies have found that learners generalize more from lessons with perceptually bland, abstract visualizations than from lessons with perceptually rich, concrete visualizations (Butcher, 2006; Goldstone & Sakamoto, 2003; Goldstone & Son, 2005; Kaminski et al., 2008, 2009; Menendez, Rosengren, et al., 2020; Petersen et al., 2014). The hypothesis is that bland, abstract representations are less tied to a specific context or scenarios, so learners are more likely to apply the information to other new scenarios, yielding better generalization.

There are reasons to doubt that abstract visualizations are always better. Recent work shows that using rich, concrete visualizations is beneficial for recall, learning, and generalization (Siler & Willows, 2014; Skulmowski & Rey, 2020; Trninic et al., 2020). Therefore, it is possible that in some situations, rich visualizations are better for learning and generalization. For example, perceptually rich visualizations might be for learning tasks that require visual retention, such as learning or mapping objects, while bland visualizations might be more beneficial for learning abstract processes (Skulmowski et al., 2021). Furthermore, rich visualizations might be more beneficial for students in early elementary school, with the benefit of bland visualizations not appearing until late elementary school or later (Menendez et al., 2022). This age-relate effect was found even though all the children were learning about the same abstract process, and thus the differences is not due to differences in the content. This suggest that the effect of perceptual richness might depend on the age of the student.

Text-picture integration

Many theories of multi-media learning propose that learners combine the information presented in text with the information presented in the visualizations to form their understanding of the lesson (Ainsworth, 2006; Mayer, 2009; Schnotz, 2014). The Cognitive Theory of Multi-media Learning (CTML, Mayer, 2009) and the Integrated Model of Text-Picture Comprehension (ITPC, Schnotz, 2014) focus on how the information from is modality is integrated. They both agree that the text and visualization are first processed separately, but the CTML proposes that learners integrate the two modalities by creating a singular mental model of the lesson that contains information from both (Mayer, 2009), while the ITPC argues that learners have two models (one for the text and one for the visualization) and integration is building connections between them (Schnotz, 2014). Empirical evidence supports the single model view (Arndt et al., 2015; Schüler, 2017; Schüler et al., 2015), suggesting that over time, learners create a single mental model of the lesson that incorporates information from both the text and the visualization.

Although both theoretical accounts posit the interaction of linguistic and visual information, most work to date has focused on manipulating only one of these factors. For example, several studies have shown how manipulating the linguistic information in a lesson can influence how people learn with a visualization (Flynn et al., 2020; Fyfe et al., 2015; Son & Goldstone, 2009). However, because the visualization is held constant, we do not know how these two features of the lesson can interact. Only a few studies have manipulated both features of the text and visualization (Arndt et al., 2015; Schüler et al., 2015). These studies showed adults general or specific texts or visuals and assess they memory for both. Their results show that after being presented with the stimuli, the information presented uniquely in the picture influences their recall of the text. After a delay, information presented uniquely in the text also influences their recall of the images (Arndt et al., 2015). Thus, supporting the view that learners create a single mental model of the lesson. These studies show that information from text and picture interact to influence recall, but they do not examine how the features of each modality can influence generalization. In the present studies, I use a similar definition of integration as the creation of a single mental model that combines information from both modalities. My focus here is not to provide support for the single model view, but rather to investigate how are visual and linguistic information typically presented together and how do different combinations influence generalization.

Current studies

In this paper, I present 8 studies examining perceptual richness and generic language in biology learning. I examine this phenomenon in the context of how people generalize facts about animals. Study 1 examines how text and visualizations are typically presented to elementary school students, by doing a content analysis of textbooks and trade books about animals. Studies

2-8 examine how the different presentations found in Study 1 might influence generalization. In Studies 2-8, participants saw generic, exemplified generic, and non-generic statements paired with either a perceptually rich or bland visualization. Even though exemplified generic statements have not been examined in prior literature on generic noun phrases, their inclusion in this study is relevant because they make reference to the visualization (see Study 1 below). Participants were asked to decide how broadly the information in the statement can be applied to the animal explicitly mentioned in the statement (e.g., chameleons) and to the broader category (e.g., reptiles). This distinction was used to examine whether category level (specific or broad) influenced generalization. Studies 2-6 use the similar method with small variations in how generalization was measured, and which factors varied within or between participants. Studies 2-7 tested undergraduate students and Study 8 tested elementary school students. For brevity, I first present Study 1, then describe the method for Studies 2-6, and an internal meta-analysis that aggregates the results of these studies. Then, I present Study 7, which eliminated the competition for visual attention between text and images (i.e., possible split attention). Finally, I present Study 8, which examines these issues with elementary school students.

Studies 2-8 were reviewed and approved by the ANONYMIZED Institutional Review Board. The hypotheses below were used for Studies 2-8 and were pre-registered for Studies 6-8. Based on prior work (Butcher, 2006; Goldstone & Sakamoto, 2003; Menendez, Rosengren, et al., 2020), I hypothesized that participants who saw the statements with bland visualizations will think that the property applies to more members of the category than those who saw the statement with a rich visualization (main effect of visualization). Based on prior research on generic language (Brandone et al., 2012; Brandone & Gelman, 2009; Tasimi et al., 2017), I also hypothesized that people would think that the statements apply more broadly when presented with generic rather than non-generic noun phrases, with exemplified generic statements falling in between (main effect of text type). I also hypothesized that this effect would be smaller when people answered for the broader category (text type by category level interaction). Finally, because no prior studies have examined the interaction between generic language and perceptual richness, I made no specific predictions about the interaction between text and visualization.

Study 1: Content analysis of elementary school science books

In order to examine if generic statements and bland visualizations were common in science educational materials, I conducted a content analysis of the visualizations and accompanying text in books aimed at elementary school students.

Book selection

Eighteen books were included in the content analysis. Four books were textbooks and 14 were trade books (see supplemental materials for details on the books). Both types of books yielded 687 images and 611 accompanying texts (1369 sentences total).

Textbooks. The textbooks were part of the MacMillan-McGraw-Hill Science Interactive Text series (2004) for grades 2, 3, 4 and 5. For each textbook, I focused on the chapters that addressed topics about animals or ecosystems. I obtained 175 images and 158 accompanying texts (259 sentences) from these books.

Trade books. I obtained the 14 trade books (i.e., books aimed at teaching children about specific concepts) from a teaching resources library. Three research assistants blind to the purpose of the study were asked to find and scan all of the images in the books in the science section of the library. They were instructed told to focus on (1) books whose primary purpose was to teach children a science concept (rather than provide a narrative), (2) books that had the preK-12 tag at the library, (3) books that had at least one picture (excluding the cover) and (4) books that addressed a topic related to animals or ecosystems. I obtained 512 images and 453 accompanying texts (1110 sentences) from these books. I obtained the intended grade information for each book from the publisher.

Coding

Coders discussed the preliminary codes and trained on coding college biology textbooks before coding the elementary school books. For the images, the coding focused on their perceptual richness. Photographs and detailed drawings were coded as rich, while diagrams, schematic drawings and drawings with fewer details were coded as bland. One coder coded all 18 books, and a second coder coded six books (one third of all books, 236 images; 35.6% of all images) to assess inter-rater reliability. Overall, reliability was satisfactory, Cohen's $\kappa = 0.78$. Other aspects of the images were also coded, see supplemental materials.

For the text, the coding focused on whether each sentence included a generic noun phrase or not. During the coding process, I found that some sentences combined generic and non-generic structures, by providing an example or directing readers to the image (e.g., "Giraffes, like the one in this picture, have long necks"). Given that previous literature has not examined these types of statements, I decided to code for them separately and will refer to these statements as *exemplified generics*. One coder coded all 18 books, and a second coder coded six books (236 accompanying texts, 40.3% of all accompanying texts; 451 sentences, 34.1% of all sentences) to assess inter-rater reliability. Reliability was lower for the texts, but still adequate (Cohen's κ = 0.65). All disagreements were resolved through discussion.

Results

I found that 90.8% (n = 624) of all visualizations were rich, and there were very few bland visualizations (9.2%, n = 63). The likelihood that a visualization was rich decreases as the intended grade of the book increases, *OR* = 0.23, $\chi^2(1, N = 686) = 79.07$, $p < .001$. When I account for non-independence due to books (i.e., the possibility that the images are not independent because images in the same book are more similar than images from different books), this relation persisted, *OR* = 0.25, $\chi^2(1, N = 18) = 13.70$, $p < .001$. This trend suggests that as children progress through elementary school, they are exposed to fewer rich visualizations, and they are exposed to more bland visualizations.

About 47.8% ($n = 655$) of the sentences that accompanied visualizations used generics and about 40.0% ($n = 548$) used non-generics (12.1% of sentences used other linguistic structures, such as questions). Because many figure captions and accompanying text include more than one sentence, I also calculated the number of captions that had at least one sentence with a generic. I found that 62.1% of the accompanying texts ($n = 427$) included at least one sentence with a generic. I also found that some of the generic sentences (2.0%, $n = 28$) were exemplified generics. I used logistic regression to predict whether the use of generics (regardless of if it was exemplified or not) depended on the intended grade of the book. However, this relation was not significant, *OR* = 1.05, $\chi^2(1, N = 610) = 0.29$, $p = .588$ (when taking into account non-independence due to books, *OR* = 1.02, $\chi^2(1, N = 18) = 0.01$, $p = .912$).

I also examined whether generic statements depended on characteristics of the images. Of the 58 bland images, 20 (34.5%) were accompanied by at least one generic statement and 38 images (65.5%) were accompanied by only non-generic statements. Of the 553 rich images, 396 (71.6%) were accompanied by at least one generic statement and 157 images (28.4%) were accompanied by only non-generic statements. I fit a logistic regression predicting the probability that the text had at least one generic. I included whether the corresponding visualization was rich and the lowest intended grade of the book as predictors. Critically, I found that rich visualizations were more likely to be accompanied by a generic than bland visualizations, *OR* = 6.28, $\chi^2(1, N = 610) = 36.90$, $p < .001$ (when accounting for non-independence due to book, *OR* = 4.85, $\chi^2(1, N = 18) = 12.22$, $p < .001$). After controlling for the characteristics of the visualizations, as intended grade level increased the probability of a sentence being generic also increased, *OR* = 1.28, $\chi^2 (1, N = 610) = 6.61$, $p = .010$, but this effect did was not significant after accounting for non-independence due to book, *OR* = 1.09, $\chi^2 (1, N = 18) = 0.18$, $p = .671$).

Discussion

This content analysis illustrates that generics and rich visualizations are common in science educational materials for elementary school students. I also found that the prevalence of these characteristics changed depending on the intended grade of the book. Similarly to findings in prior work, all books included generic statements (Gelman et al., 2013), but these statements might be more common in books for older students. Notably, I also identified a new type of generic statement, exemplified generics, which include a clause with an example of the category and tend to refer to the visualization. Given that prior work has not examine how students interpret exemplified generics, I explore this issue in the subsequent studies.

Rich visualizations were very common in all books, but their proportion decreased depending on the intended grade, with fewer rich visualizations in books targeted at older students. This trend parallels previous content analyses of science textbooks for middle school, high school, and college students (Wiley et al., 2017), which similarly find that the proportion of rich decorative visualizations decreases with grade and that the proportion of abstract, diagrammatic bland visualization increases with grade. Therefore, suggesting that as students advance in grades they are exposed to fewer rich visualization and more bland visualizations. If true, this change in the visualizations in children's environments could explain why previous work has found that rich visualizations are useful for children in early elementary school to learn biological topics, but older elementary school students and undergraduate students generalize more broadly with bland visualizations in the same tasks (Menendez et al., 2022; Menendez, Rosengren, et al., 2020). This content analysis also shows that generics often accompany visualizations, and their prevalence seems to depend on characteristics of the visualization. Therefore, it is important to investigate how learners make sense of the combination of visual and linguistic information, particularly for mismatch cases in which the visualization is rich (which does *not* promote generalization) but the language is generic (which promotes generalization).

Studies 2-6: Undergraduate students

I conducted five studies examining how undergraduate students generalize animal facts. All studies manipulated the text type (generic, exemplified generic, non-generic), and visualization type (bland, rich). Additionally, to examine how far students generalized, I also manipulated whether the question asked them about the category mentioned in the lesson, or a broader category (I will refer to this as category level in the remaining of the manuscript). Complete demographic information for the participants in Studies 2-6 can be found in the supplemental materials.

Method

Participants

In Study 2, I recruited 116 undergraduate students (*M* = 18.55 years, *SD* = 0.77 years; 59 women, 56 men, and 1 did not respond; 76.7% white). Of these 116, 96 students passed all attention checks. In Study 3, I recruited 112 undergraduate students (*M* = 18.67 years, *SD* = 1.38 years; 60 women, and 52 men; 68.7% white). Of these 112, 99 students passed all attention checks. In Study 4, I recruited 139 undergraduate students (*M* = 18.59 years, *SD* = 0.81 years; 85 women, 51 men, and 1 did not respond; 69% white). Of these 139, 110 students passed all attention checks. In Study 5, I recruited 90 undergraduate students (*M* = 18.73 years, *SD* = 1.24 years; 59 women, 28 men, and 2 did not respond; 63.3% white). Of these 90, 51 students passed all attention checks.

In Study 6, I recruited 300 undergraduate students enrolled in an Introduction to Psychology course to participate in this study for extra credit. I determined this sample size based on the effect size found in Study 5 of *OR* = 0.87 for the effect of visualization (as this is methodologically the most similar study to Study 6 and this effect is smaller than the effect of text). I used this value in power analysis using https://jakewestfall.shinyapps.io/crossedpower/ participant with condition design. Details on the inputted values can be found in the supplemental materials. According to the power analysis, a sample size of 218 would allow me to detect the predicted effects with 80% power. I increased the sample size to 300 to account for the exclusion of those who might fail the attention checks.

I recruited 300 undergraduate students (*M* = 18.38 years, *SD* = 0.81 years; 199 women, 100 men and 1 did not respond; 65% white). Of these 300 participants, 152 passed all the attention checks.

In all studies, participants who did not pass all of the attention checks were excluded from the analyses.

Materials

I included 28 animals in the study. I collected cartoon-style images with white backgrounds of all the animals from Google images. All of these images were colorful drawings that included shading and textural details, and these were the rich images used in the studies. I modified the rich images to create a perceptually bland version that was only a line drawing of the animal. For Studies 2-5, the bland images were black and white versions of the rich images. For Study 6, the bland images were made blander by removing additional shading and textural details such that they resembled outlines of the rich images. Each animal was associated with one property, 21 with structural properties (e.g., "have tetrodotoxins inside," or "have *sticky* spit") and 7 with behavioral properties (e.g., "prey on antelopes," see supplemental materials for a list of the animals and properties). None of the properties selected were displayed in the visualization. All of the properties mentioned were true for at least some of the members of the category.

Each statement could be presented as a generic (e.g., "Crocodiles have two atria"), an exemplified generic (e.g., "Crocodiles, like the one in the picture, have two atria"), or a non-generic phrase (e.g., "This crocodile has two atria"). The picture was always presented above the statement. After reading the statement, participants were asked to indicate how many other category members would also have the property (e.g., crocodiles). Then, participants were asked the same question about members of the broader category level (e.g., reptiles). In Studies 2-4, participants were asked "what percentage of [animals] do you think [statement]?" and then entered a number between 0% and 100% as their answer. In Studies 5-6, participants were asked "How many other [animals] do you think [statement]?" and they could say that "None," "A few," "Some," "Most," or "All" other category members would share the properties. Participants' responses to these questions are the dependent variable. Participants saw this question twice, first with the animal that was mentioned in the statement (specific category), and then with the broader category (the taxonomical *class* of the organism depicted). This difference in the category mentioned in the question comprises the manipulation of category level, and was done within subjects in all studies.

To eliminate the possibility that participants did not know which animals belonged to which class, in Study 6, before every statement describing the feature of interest, there was a generic statement explaining the relation between the specific and broader category (e.g., “Crocodiles are reptiles. This crocodile has two atria.”). This meant that, for Study 6, when participants saw non-generic statements, participants actually saw a generic statement followed by a non-generic statement describing the property of interest. As shown in the content analysis, having generic and non-generic statements captioning an image is common in science books. See Figure 1 for an example of the stimuli.

Procedure

Participants completed the study online using the Qualtrics system. After consenting to participate in the study, participants read the instructions, which told them that they were going to be shown a picture of an animal accompanied by a statement, and that they would have to answer two questions based on the image and statement. I specifically instructed participants not to search online for any of the animals or statements presented during the study. After participants read the instructions, they were presented with the primary task. One animal (image, statement, and questions) was presented per page. The presentation of the image and statements was self-paced, meaning that participants could spend as much time as they wanted in any given animal. Interspersed throughout the study were attention checks. These attention checks showed the image of an animal and a statement similar to an actual trial, but in the statement or the question I told the participant to select a particular response (e.g., “Please select ‘Some’”). The same attention checks were used in all conditions (but the image and text type were modified to match participant's random assignment).

After completing all items, participants provided demographic information, including gender, race, ethnicity, year in school, and whether they had searched for any of the statements during the study. Finally, participants were debriefed as to the purpose of the study.

Design

I used a 3 (Text type: Generic, Exemplified generic, Non-generic) x 2 (Visualization type: Rich, Bland) x 2 (Category level: Specific, Broader) design. Participants saw 28 animals. In *Study 2,* text type, visualization type and category level varied within subjects. In *Study 3,* visualization type and category level were varied within subjects and text type varied between subjects. In *Study 4* text type and category level varied within subjects and visualization type varied between subjects. In *Study 5,* text type and category level varied within subjects and visualization type varied between subjects. In *Study 6,* text type and category level varied within subjects and visualization type varied between subjects.

When the factor varied within subject, each animal for each participant was randomly assigned to one of the conditions. Participants always saw the questions for the specific and broader category. The Studies 2-4 asked participants "what percentage of [animals] do you think [statement]?" Participants then entered a number between 0% and 100% as their answer. Study 5-6 asked "how many other [animals]…" and participants responded by clicking one of the five categories (see details in method section below).

Data analytic approach

Rather than present the results of each study separately, given the similarity in their design I conducted an internal meta-analysis (also called a mini-meta-analysis; Goh et al., 2016) of Studies 2-6. This technique allows me aggregate across the studies and establish the robustness of the findings. Separate analyses for each study can be found in the supplemental materials. Combined, these studies included 757 participants, with 508 that passed all the attention checks.

Because some of the variables varied within or between subjects in different studies, it is not recommended to use raw means to examine the effect of each factor, necessitating that I use pre-calculated effect sizes (Harrer et al., 2021), see supplemental materials. I meta-analyzed every estimate of the five studies using fixed effects in which the mean Cohen's *d* was weighted by the size of the confidence interval. Because I used this approach for every estimate, there were in total 15 separate meta-analyses. To perform these analyses I used the *meta* (Balduzzi et al., 2019) and *metafor* (Viechtbauer, 2010) packages in R. To asses between study heterogeneity, I used I^2 (Higgins & Thompson, 2002). I^2 values < 25% suggest low heterogeneity, 25-50% suggest moderate heterogeneity, and 50-75% suggest substantial heterogeneity.

Transparency and Openness

I report all data exclusions, manipulated variables, conditions and outcome variables. Additional coding categories for Study 1 are reported in the supplemental materials. Studies 2-5 were not pre-registered and Studies 6-8 were pre-registered (Study 6 see https://osf.io/cm3w9/?view_only=c8e367e945b54fa9bde937fe6920976a; Study 7 see https://osf.io/rbguz?view_only=c394c55989d84774a47ef1b21be6c0a3; Study 8 see https://osf.io/5trbc/?view_only=4af865c07bfb48c7bb46f4bbedbbcd8e). All data, analysis scripts, and materials associated with Studies 1-8 can be found at: https://osf.io/d9zmq/?view_only=50cb6a81311c4f98adf1db6bcaa9bd0d.

Results

Overall, there was a large difference between generic and non-generic statements, *d* = 0.857 [0.712, 1.002], with low between-study heterogeneity, I^2 = 2.1%. There was also a small difference between generic and exemplified generic statements, *d* = 0.265 [0.144, 0.387], with low between-study heterogeneity, I^2 = 0.0%. There was a medium-sized difference between exemplified generics and non-generic statements, *d* = 0.580 [0.446, 0.715], with low between-study heterogeneity, I^2 = 21.0%. These effects show that participants generalized most when they saw a generic, followed by exemplified generics and then non-generics. See Figure 2.

I also found a reliable effect of category level, *d* = 1.242 [1.113, 1.370], but there was substantial heterogeneity, I^2 = 91.2%. This overall suggest that participants generalized more to the specific category than the broader category. I also found reliable interactions between category level and the generic vs. non-generic contrast, *d* = 0.811 [0.651, 0.972], I^2 = 0.2%, between category level and the exemplified generic vs. non-generic contrast, *d* = 0.609 [0.453, 0.765], I^2 = 0.0%, and between category level and the generic vs. exemplified generic contrast, *d* = 0.168 [0.003, 0.333], I^2 = 0.0%. As can be seen in Figure 3, the difference in statement type where present for the specific category but not the broader category.

There was no effect of the perceptual richness of the visualizations, *d* = 0.08 [-0.04, 0.201], with low between-study heterogeneity, I^2 = 0.0%. No other effects were significant. Figures and results of these meta-analytic averages are available in the supplemental materials.

Discussion

Overall, the findings of the internal meta-analysis suggest that there is strong evidence that text type influences generalization, with generic statements promoting generalization the most, then exemplified generics, and then non-generics. Critically, the difference between generics and exemplified generics was not significant in every study, but the meta-analytic average was reliably different from zero. Undergraduates were also more likely to extend the statement to other members of the same category than to members of the broader category. I also found that across studies, these differences between the types of text decrease when participants are making inferences about the broader category. The meta-analytic analysis shows that these results are robust across samples, outcome measures, whether participants read a generic statement indicating the relation between the broader and specific category, and whether the factors vary within or between participants. Importantly, I did not find any reliable effects of visualization type or any significant interactions that involved visualization.

One possibility is that by having both the picture and text on screen, I created competition for visual attention (i.e., split attention). Students might have had to choose whether the look at the picture or the text. To examine if this was the case, in Study 7, I eliminated this competition for visual attention by creating videos for all the statements and answer choices. Although the text was still present in the screen, the fact that it was read out loud for participants means that they could have listened to the statements and devoted their visual attention to the visualizations. Additionally, I analyze the data using a Bayesian framework. This allows me to examine whether the effect of visualization includes 0 and what the distribution of likely effects is.

Study 7: Undergraduate students

Method

Participants

I recruited 300 undergraduate students (*M* = 18.49 years, *SD* = 0.95 years; 177 women, 113 men, 2 non-binary, and 8 did not respond) enrolled in an Introduction to Psychology course to participate in this study for extra credit. The racial and ethnic breakdown of the sample was: 58.7% White (*n* = 176), 3.0% Black (*n* = 9), 21.3% Asian (*n* = 64), 4.0% Hispanic or Latinx (*n* =

12), 1.0% Middle eastern (n = 3), 0.33% Native American (n = 1), and 8.7% bi- or multi-racial (*n* = 26); 2.7% did not report race or ethnicity information (*n* = 8). Most of the participants were native speakers of English either exclusively (*n* = 213), or in addition to another language (*n* = 46), but 25 participants were not native speakers of English (15 did not respond). This sample size was based on the same power analysis used for Study 6, using the effect size for the effect of visualization type found in Study 5. Details on the power analysis can be found in the supplemental materials but the minimum number of participants to detect the effect with 80% was 218, and I oversample to account exclusions due to the attention checks. However, 154 participants did not pass all the attention checks and were excluded from the analyses, leaving a final sample of 146 participants.

Design

The design of this study is identical to that of Study 6, except that instead of viewing images, participants watched videos in which a narrator read the statement and questions out loud. The video of the stimuli played automatically, and the stimuli advanced automatically in order to make sure all participants saw the stimuli. This meant that rather than being a self-paced study, as with Studies 2-6, participants had a limited time to process the statements.

Materials & Procedures

Instead of having participants read the statements, the survey automatically played a video that showed the animal and the statement, and a narrator read the statement, questions and response options out loud. Participants answered the question “How many other [animals] do you think [statement]?” by saying “None,” “A few”, “Some,” “Most,” or “All.”

Data analytic strategy

Analyses were performed under a Bayesian framework, first fitting a baseline linear model with skeptical priors and then fitting increasingly complex models. I used leave-one-out cross-validation and used the difference in *elpd* to select the model that best fit the data. For all models, I set non-generic as the reference category and used non-orthogonal contrasts to examine the effect of text type. More details on the Bayesian data analytic technique can be found in the supplemental materials. For each best fitting model, I report the beta (the median of the posterior distribution in log odds), and 95% highest density intervals (HDI). If the HDI includes 0, then I calculate the probability of direction (i.e., the percentage of the posterior distribution that is in the same direction as the beta) as a way to determine *where* zero falls in the distribution. For conciseness, I report the probability of direction only when the value is higher than 85%. I used a Bayesian approach for Studies 7-8 because they allow me to examine if the effect of a variable is indeed 0, and thus claim whether the variable has no effect. Parallel analyses under a frequentist framework can be found in the supplemental materials alongside analyses using ordinal logistic regression (rather than linear regression). All means and standard deviations reported in the text are unadjusted for the other variables in the model (e.g., the mean for the different statement types is not adjusted for visualization type).

Results

Pre-registered analysis

The best fitting model included the interaction of text type and visualization type and the interaction of text type and category level. See Table 1. Replicating the previous studies, participants answered that the statement applied more broadly if they heard (or heard and read) a generic (M = 3.72, SD = 1.25) rather than a non-generic statement (M = 3.57, SD = 1.22), b = 0.33 [0.26, 0.40]. Additionally, participants answered that the statement applied more broadly when they heard an exemplified generic (*M* = 3.41, *SD* = 1.19) rather than a non-generic statement, *b* = 0.16 [0.09, 0.23]. I recentered the model to examine the difference between generic and exemplified generic statements, and found evidence that participants answered that the statement applied more broadly if they heard a generic rather than an exemplified generic statement, *b* = 0.17 [0.12, 0.23].

I also found evidence of an effect of category level, such that participants thought that the statement applied more broadly to the category mentioned in the statement (*M* = 4.31, *SD* = 0.93) than to the broader category (*M* = 2.83, *SD* = 1.03), *b* = -1.25 [-1.40, -1.09]. There was also evidence for interactions between text type and category level, Generic vs Non-generic: *b* = -0.42 [-0.52, -0.32], and Exemplified vs Non-generic: *b* = -0.19 [-0.29, -0.08]. As can be seen in Figure 4, the differences between text types were present only when reasoning about the specific category mentioned in the statement. There were no differences among the text types for the broader category.

In contrast with the previous studies, but in line with the pre-registered hypothesis, there was evidence for an effect of visualization type, *b* = -0.16 [-0.29, -0.02], such that undergraduate students generalized more broadly when the visualization was bland (*M* = 3.61, *SD* = 1.24) than when it was rich (*M* = 3.53, *SD* = 1.22). There was also an interaction between visualization type and text type, Generic vs Nongeneric: *b* = 0.17 [0.03, 0.31], and Exemplified vs Nongeneric: *b* = 0.09 [-0.05, 0.23], probability of direction 89.57%, suggesting that the greater generalization with the bland visualization occurs only when the statement does not use generic language. See Figure 5.

Discussion

Study 7 shows that undergraduate students do attend to characteristics of both the text and visualization when generalizing. The results of the study support the initial hypotheses (and findings of the internal meta-analysis) that students will generalize more when the text is generic rather than non-generic. In line with the initial hypothesis (but different from the internal meta-analysis), they also generalize more with bland visualizations. This replicates prior findings with undergraduate students on other biological (Menendez, Rosengren, et al., 2020) and non-biological topics (Goldstone & Sakamoto, 2003; Kaminski et al., 2008). The difference in the results from Studies 2-6 to Study 7 are likely due to difference in visual attention. Although the text was also present in this study, participants did not need to read because there was a voice over. Therefore, participants could devote their visual attention to the picture rather than examine only the text. This aligns with previous cognitive models of multimedia learning that highlight the importance of visual attention on how people learn with visualizations (Johnson & Mayer, 2012; Mayer, 2008, 2009). However, the difference could also be due to participants having a limited amount of time to process the statements. Interestingly, there was also an interaction between generics and perceptual richness. Undergraduate students seemed to generalize more broadly when at least one element of the lesson promoted generalization. If the text was generic, then they generalized broadly. If the text was not generic but the image was bland then they generalized broadly. This suggest that, although each element promotes generalization, there does not seem to be an added benefit of having multiple elements that promote generalization.

Although the findings with undergraduate students are interesting, it is critical to understand how children integrate visual and linguistic information. As a reminder, the content analysis presented earlier involved elementary school science books, making this a theoretical and practical question for children's science learning. Additionally, work on visualizations has suggested that there is a developmental change in how children learn with visualizations in science (Menendez et al., 2022). Children in early elementary school seem to learn better with rich visualizations, while children in late elementary school seem to generalize more with bland visualizations. However, in that study, children received one lesson about metamorphosis, and then were asked about life cycle changes of other animals. The current study is much simpler, in that it presents statements to children and then asks them how broadly the statements apply. Therefore, it is of interest to examine whether the results of Menendez et al. (2022) extend to simpler tasks. To do this, I recruited 6- to 9-year-old children to participate in this study. This age group is comparable to the age groups in previous studies of children's understanding of visualizations (Menendez et al., 2022) and generics (Leshin et al., 2021). In the analysis of this study, I also examine whether grade (as a continuous variable) influences children's generalizations.

Study 8: Elementary school children

Method

Participants

I recruited 108 first to fifth grade children to participate in this study (*M* = 9.11 years, *SD* = 1.26 years, Range: 6-12; 60 girls, 47 boys, and 1 non-binary individual). Families were recruited through a database of families that had previously participated in studies in a psychology research lab, through postings on social media, and via word of mouth. Participants received a $10 gift card for participating in this online study. I pre-registered that I would collect data from 100 children (a similar sample size as prior studies with children in this area, Menendez et al., 2022; Muradoglu et al., 2022), however due to the unmoderated nature of the study more children completed the study than intended. As specified in the pre-registration, all families were recruited before April 30th, 2022.

The racial and ethnic breakdown of the sample, according to parental reports, was: 73.1% White (*n* = 79), 2.8% Black (*n* = 3), 4.6% Asian (*n* = 5), 2.8% Hispanic or Latinx (*n* = 3), and 14.8% bi- or multi-racial (*n* = 16);1.8% did not report race or ethnicity information (*n* = 2). Most of the participants were native speakers of English either exclusively (*n* = 98), or in addition to another language (*n* = 6), but 4 participants were not native speakers of English.

Materials & Procedures

All materials, design and procedures were identical to Study 7. To make it easier for children to complete, I used an unmoderated study procedure (Rhodes et al., 2020). Families received a link to participate in the study and a Zoom link for them to record their child completing the study (with instructions for how to do so). Once families opened the Zoom link, the session was automatically recorded. First, parents were asked to provide informed consent. Then, children provided verbal assent. Video instructions explained the task to the children. Parents were told that they could stay in the room with the child and help them to navigate the study or click their answers, but they were explicitly asked to not provide the answers to their child. Children completed the study in their own time and for as long as they wanted. At the end of the session, the video and transcript of the session was automatically sent to the research team.

Results

Pre-registered analyses

The best fitting model included an interaction between text type and visualization type, and an interaction between text type and category level. See Table 2. As in Studies 2-7, there was evidence for an effect of text type. Children generalized to other category members more when the statement included a generic (*M* = 3.48, *SD* = 1.31) than an exemplified generic (*M* = 3.42, *SD* = 1.27), *b* = 0.06 [-0.01, 0.12], probability of direction = 96.12%, or a non-generic (*M* = 3.35, *SD* = 1.30), *b* = 0.13 [0.06, 0.19]. Children also generalized more when the statement included an exemplified rather than a non-generic, *b* = 0.07 [-0.01, 0.14], probability of direction = 96.24%. There was also evidence for an effect of category level, such that children generalized more to the specific category mentioned in the statement (*M* = 4.13, *SD* = 1.08) than to the broader category (*M* = 2.71, *SD* = 1.09), *b* = -1.31 [-1.48, -1.14]. As with the adults, the difference between generic and non-generic statements disappeared when children made inferences about the broader category, *b* = -0.16 [-0.28, -0.04]. The difference between generic and exemplified generic also disappeared when children made inferences about the broader category, *b* = -0.15 [-0.27, -0.02]. However, there was little evidence for an interaction between exemplified generic vs. non-generic and category level, *b* = -0.03 [-0.15, -0.10]. See Figure 4.

There was also evidence for an interaction between generic v. non-generic and visualization type, *b* = 0.13 [-0.01, 0.26], probability of direction 96.76%, between exemplified generic v. non-generic and visualization type, *b* = 0.09 [-0.04, 0.22], but not between exemplified generic vs. non-generic and visualization type, *b* = 0.04 [-0.09, 0.18]. As can be seen in Figure 5, there was no difference between the rich and bland visualizations when the text had a non-generic or an exemplified generic, but when the text was generic, children generalized more with the rich visualization than with the bland visualization. Finally, there was also evidence for an effect of grade, such that older children generalized more than younger children, *b* = 0.09 [0.03, 0.14].

Discussion

Overall, the results of Study 8 show that children generalize more with generic statements, than exemplified generic statements, and finally non-generic statements. As in Studies 2-7, this pattern held only for the specific category mentioned in the statement, and these differences disappeared for inferences about the broader category. Additionally, with age children generalized more to the specific category. Critically, I found that the richness of the visualization interacted with the text to influence children's generalization. Children were more likely to generalize when the text was generic, but the visualization was rich.

General discussion

Studies 2-8 revealed some similarities between children and adults. First, there was consistently an effect of text type, such that both children and adults generalized more with generic statements, followed by exemplified generic statements, and followed by non-generic statements. Second, the effects of text type diminished when participants made inferences about the broader category. This suggests that characteristics of the text influence generalization similarly across elementary school and beyond. It is worth noting that there were also developmental differences, such that with age, children came to generalize more for the specific category mentioned in the statement. The studies also show developmental differences in how people use visualizations to make inferences. The internal meta-analysis shows that adults were not influenced by the richness of the visualization, and Study 7 shows that went the competition for visual attention is reduced, adults generalize more broadly with the bland images. In Study 8, however, I found that children were influenced by perceptual richness, such that they generalized more to the specific category when the visualization was rich and the text was a generic.

Generalizing from text

All the studies show consistent findings in how students generalize from text. The use of generic noun phrases was consistently associated with greater generalization in all studies, showing that this effect is robust across age groups, outcome measures, and methodological decisions (i.e., whether it varied within or between participants). These studies also identified exemplified generics as a subtype of generics that are sometimes used in science books, and I showed that students treat these statements differently than generics and non-generics, leading to levels of generalization in between those of generics and non-generics. Finally, I also showed that students were reluctant to generalize to the broader category, showing the difficulties of generalizing to superordinate categories (Lopez et al., 1992; Osherson et al., 1990).

There were also developmental differences in the effect of generics. First graders were somewhat conservative with their generalization of rare animal facts, but older children generalized more broadly. Additionally, adults showed greater differences between the statement types, suggesting that over development, students learn to differentiate between different linguistic features. This developmental difference could be reflective of differences in language development, however, past work shows that even children younger than those included in Study 8 understand and use generics (Gelman et al., 1998, 2013). This suggests that these differences are not due solely to language development. Rather, it could be due to developmental differences in inductive inference. The task in these studies could be construed as an inductive inference task, in which participants learn about a property of one or at least some (depending on the text type) category members and are asked to make inferences about other category members. Prior work on inductive inference has found that, with age, children are more likely to make these inferences (Sloutsky et al., 2001). Therefore, this age differences could be due to differences in inductive reasoning.

Generalizing from visualizations

Studies 7 and 8 showed that students do use the richness of visualizations to guide their generalization, however, this too changes with age. The shift from richness being useful for elementary school students, at least when paired with a generic statement, to constraining generalization for undergraduate students suggest a shift in how students interpret visualizations. Prior work by Menendez et al. (2022) using a different task showed a developmental trend in which the benefit of bland visualizations emerges slowly over the elementary school years. This trend suggests that the advantage for bland and abstract visualizations that was seen in prior work (Goldstone & Sakamoto, 2003; Kaminski et al., 2008, 2013; Kaminski & Sloutsky, 2013) is not universal. Rather, this advantage seems to be context dependent (De Bock et al., 2011; Siler & Willows, 2014; Skulmowski, 2022), varying with the characteristics of the task and learner.

The age differences in how students use perceptual richness to generalize might be related to the trends found in Study 1. Study 1 showed that elementary school science books typically have very detailed visualizations such as photographs. Books intended for students in later grades contained fewer rich images, a trend that continues into college (Wiley et al., 2017). This could potentially mean that students perform best with the types of representations they are used to seeing in their environment. Elementary school students, who are typically exposed to rich visualizations, generalize more when they saw rich visualizations. Furthermore, they generalized most when the text was generic and the visualization was rich, the most common combination of text and image found in elementary school science books. Undergraduate students are more commonly exposed to bland visualizations (Wiley et al., 2017), and also generalized more with them. This might indicate that using visualizations with features that students typically encounter might be more beneficial for them.

It is important to acknowledge that there are differences in the content of the books aimed at different grades, which could explain the differences in visualizations. While many of the concepts in the elementary books could easily be depicted visually (e.g., topics like animals, fossils, or different ecosystems), this is not true of every concept. Particularly at higher grade levels, students might be learning topics that cannot be photographed or easily depicted (e.g., evolution, or the Krebs' cycle). Thus, the need for more schematic or diagrammatic illustrations in books aimed at students in higher grades. This is not a confound in the present experiments as the content was the same across age groups, but could explain why there is a difference in the science books. However, even scientific diagrams can be perceptually rich, as has been shown with life cycle diagrams (Menendez, Mathiaparanam, et al., 2020) and pedigree diagrams (Mathiaparanam et al., 2022). Therefore, the effect of perceptual richness might still be relevant at all grade levels.

Text-picture integration

The studies presented here are also among the few studies that manipulate *both* the verbal and visual information in a lesson and measure learning or generalization. Although several researchers have proposed theories of how people integrate visual and verbal information (Ainsworth, 2006; Mayer, 2008; Moreno, 2007; Schnotz, 2014), the majority of the studies only manipulate one factor. Studies that manipulate both factors have found that students create a single coherent model that combines information from both modalities (Arndt et al., 2015; Schüler et al., 2015), and that this influences what information students recall. I found that children use cues from both the text and the visualization to infer the generality of the information presented—generalizing most when the text suggests the statement is broadly applicable, but the visualizations are rich. Meanwhile, undergraduate students seem to rely primarily on the text, but when split attention is reduced they generalize broadly when either the text or image promotes generalization. Undergraduate students in Studies 2-6 likely ignored the visualizations due to competition in visual attention. There is work suggesting that people sometimes ignore visualizations (Keehner et al., 2008; Stull et al., 2012), particularly when they believe they are not relevant (Eitel et al., 2019). Therefore, undergraduate students might have defaulted to ignoring the visualization and relying primarily on the text unless they had the resources to attend to both.

The difference in how adults and children use visual and verbal information to inform their generalization can also be related to work on inductive inference. Prior work has shown that adults rely primarily on labels (i.e., verbal information) and children consider both labels and perceptual similarity (i.e., visual information; Sloutsky & Fisher, 2004). Therefore, there seem to be similar developmental differences in the strategies that people use to integrate text and visualization to guide generalizaitons. This suggests the need for future research that examines when and why people decide to attend to visualizations and how that might vary with age.

Whether the results of these studies are due to learners generating a single coherent model (Mayer, 2009) or connecting two models from each modality (Schnotz, 2014) is unclear from the results of these studies. Participants could complete the task without creating a single model, or even without attending to the visualization at all! However, this was not the main goal of these studies. Rather the focus was on how participants use both types of information when making generalizations. If adults are using a single model, then it is possible that indicating either visually or linguistically that the information is broader makes this mental model more general. If they are building connections between models, it is possible that they used the information in the different modalities to constrain or broader their interpretation of the information in the other.

Cues to generality hypothesis

Beyond simply stating that students attend to both text and images, it is also important to understand how they are using the two to guide their generalizations. Generalization can be construed as an inference problem; students have to make inferences about how far the information provided in the lesson extends. To solve this inference problem, students might attend to the features of the different elements of the lesson. Features like perceptual richness or generics might serve as cues for students that they should generalize more broadly. Other characteristics of a lesson (beyond just linguist and visual information) could also serve as cues for students to decide how broadly to generalize. Critically, this proposal focuses on student using these characteristics to guide their generalization. If students are unfamiliar or lack experience with a particular feature (as might be the case for early elementary school students learning with bland visualizations), then they might ignore that feature when making generalizations. As such, the importance is not on the features themselves, but on the inferred communicative intend of the features.

This perspective could also explain why there is a difference between the generic and exemplified generic statements. Linguistically, these statements were identical, except that the exemplified generic included a clause that made reference to the visualization. If using only the linguistic information, this clause should not matter. If you interpret “Chameleons have sticky spit” as meaning that all chameleons have sticky spit, then being told about an example of a category member that shows that property should *not* constrain your generalizations. However, I found a decrease in generalization for the exemplified generic, relative to the generic, for all age groups and most clearly in the adults. This difference was also present in Study 3, were participants saw only one type of noun phrase, and so the difference are likely not due demand characteristics pushing participants to select different answers for different noun phrases. Participants might have thought that, pragmatically, there is a difference between the generic and exemplified generics. They might have inferred that the information must be more relevant to the specific type of the animal shown, because otherwise, there would be no reason for the speaker to qualify it. This would suggest that people do not use only the linguistic information in the statement, but they also reason pragmatically about the statements.

This proposal could also explain the different findings between elementary school and undergraduate students. Undergraduate students generalized more broadly when the lesson included at least one cue to generality. When the text included a generic, they generalized broadly. In the absence of a generic, they examined the image and if it was bland, then they generalized broadly. This suggests that adults were attending to the features of both the text and the image to decide how broadly to generalize. Elementary school children also focused on features of the text. However, because they might have little experience with perceptually bland images, they might not have interpreted the lack of details as meaning that they should generalize broadly. Thus, in the absence of generics, they generalized similarly with both bland and rich visualizations. This proposal, however, does not explain why children generalized more broadly when the text was generic and the image was rich. Although it could be due to children's experience with this combination of features, more work needs to examine how children use perceptual richness as a cue in their generalizations.

Limitations

It is important to acknowledge the limitations of these studies. First, there was a high percentage of undergraduate participants who did not pass the attention checks. Although the reduction in power is partially mitigated by the internal meta-analysis, the high exclusion rates might also suggest that the results might not be representative of all students. Second, although the difference in results between Study 2-6 and Study 7 could be attributed to visual attention, this is not certain as I did not have a measure of visual attention. Replications of these studies with eye-tracking measures might be necessary in order to determine whether adults are choosing to ignore the visualizations altogether when they have to read the text. Third, the images used in the Studies 2-8 are not an accurate representation of the visualizations students encounter. The majority of the rich visualizations found in Study 1 were detailed photographs, rather than the drawings used in the experiments. Similarly, the bland images tended to be diagrams rather than the line drawings used in the experiments. Future research should examine how differences between photographs and drawings or line drawings and schematic visuals influence learning and generalization.

Finally, and perhaps most importantly, the task in this study was relatively simple, and though analogous, it is not representative of the complexity of text and visual information that people encounter when reading a science book. Although the underlying research question (how do the characteristics of the text and images used in lessons influence learning and generalization) has clear educational implications, the task in these studies were very simple, the assessments were not comprehensive, and the task was repetitive. In very few educational environments students will be asked to read two-sentence statements with a picture and then immediately answers questions about what they just read, nor will adults and elementary school students receive the exact same lesson. The lack of external validity of these studies was necessary in order to investigate whether the linguistic and visual features of lessons *can* affect learning and generalization. If no effects were found in this simple task, the likelihood that these features influence learning and generalization in actual educational settings would be low. However, the fact that these features had effects in this setting does not mean that these effects will extend to real world educational settings. Future studies should examine these issues with more complex visualizations, like photographs or scientific diagrams, or with text that is longer. These conditions might more faithfully emulate the conditions people typically encounter in outside the lab, and might provide more evidence before extending the findings to classrooms or other educational settings.

Conclusions

Overall, these studies suggest that both children and adults use linguistic cues to infer the generality of statements conveyed in multimedia lessons. I found that people applied statements broadly when they had a generic noun phrase, were more constrained when the generic noun phrase was followed by a clause referencing the image (i.e., exemplified generic), and generalized least when the statements did not include a generic noun phrase. People also made different inferences depending on whether they had to think about the specific category mentioned in the statement or the broader category. We also found developmental differences in the use of visualizations. Children generalized most when the text was generic and the image was rich, while undergraduate students generalized when either the text was generic or the image was bland. This suggests that both text and visualizations influence inferences from multimedia lessons but have different effects, depending on the age of the learner.

References

Ainsworth, S. (2006). DeFT: A conceptual framework for considering learning with multiple representations. *Learning and Instruction*, *16*(3), 183–198. https://doi.org/10.1016/j.learninstruc.2006.03.001

Arndt, J., Schüler, A., & Scheiter, K. (2015). Text–picture integration: How delayed testing moderates recognition of pictorial information in multimedia learning. *Applied Cognitive Psychology*, *29*(5), 702–712. https://doi.org/10.1002/acp.3154

Balduzzi, S., Rücker, G., & Schwarzer, G. (2019). How to perform a meta-analysis with R: A practical tutorial. *Evidence Based Mental Health*, *22*(4), 153–160. https://doi.org/10.1136/ebmental-2019-300117

Brandone, A. C., Cimpian, A., Leslie, S.-J., & Gelman, S. A. (2012). Do lions have manes? For children, generics are about kinds rather than quantities. *Child Development*, no-no. https://doi.org/10.1111/j.1467-8624.2011.01708.x

Brandone, A. C., & Gelman, S. A. (2009). Differences in preschoolers' and adults' use of generics about novel animals and artifacts: A window onto a conceptual divide. *Cognition*, *110*(1), 1–22. https://doi.org/10.1016/j.cognition.2008.08.005

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

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

Cimpian, A., & Markman, E. M. (2009). Information learned from generic language becomes central to children's biological concepts: Evidence from their open-ended explanations. *Cognition*, *113*(1), 14–25. https://doi.org/10.1016/j.cognition.2009.07.004

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.

Eitel, A., Bender, L., & Renkl, A. (2019). Are seductive details seductive only when you think they are relevant? An experimental test of the moderating role of perceived relevance. *Applied Cognitive Psychology*, *33*(1), 20–30. https://doi.org/10.1002/acp.3479

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

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

Garner, R. (1992). Learning from school texts. *Educational Psychologist*, *27*(1), 53–63.

Garner, R., Alexander, P. A., Gillingham, M. G., Kulikowich, J. M., & Brown, R. (1991). Interest and learning from text. *American Educational Research Journal*, *28*(3), 643–659.

Garner, R., Gillingham, M. G., & White, C. . S. (1989). Effects of "seductive details" on macroprocessing and microprocessing in adults and children. *Cognition and Instruction*, *6*(1), 41–57. https://doi.org/10.1207/s1532690xci0601_2

Gelman, S. A., Coley, J. D., Rosengren, K. S., Hartman, E., & Pappas, A. (1998). Beyond labeling: The role of maternal input in the acquisition of richly structured categories. *Monographs of the Society for Research in Child Development*, *63*(1), i–157. https://doi.org/10.2307/1166211

Gelman, S. A., Ware, E. A., Manczak, E. M., & Graham, S. A. (2013). Children's sensitivity to the knowledge expressed in pedagogical and non-pedagogical contexts. *Developmental Psychology*, *49*(3), 491–504. https://doi.org/10.1037/a0027901

Goh, J. X., Hall, J. A., & Rosenthal, R. (2016). Mini meta-analysis of your own studies: Some arguments on why and a primer on how: mini meta-analysis. *Social and Personality Psychology Compass*, *10*(10), 535–549. https://doi.org/10.1111/spc3.12267

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

Harrer, M., Cuijpers, P., Furukawa, T. A., & Ebert, D. D. (2021). *Doing meta-analysis with r: A hands-on guide*. Chapman & Hall/CRC Press. https://bookdown.org/MathiasHarrer/Doing_Meta_Analysis_in_R/pooling-es.html#binary-outcomes

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, *21*(11), 1539–1558. https://doi.org/10.1002/sim.1186

Hoicka, E., Saul, J., Prouten, E., Whitehead, L., & Sterken, R. (2021). Language signaling high proportions and generics lead to generalizing, but not essentializing, for novel social kinds. *Cognitive Science*, *45*(11), e13051. https://doi.org/10.1111/cogs.13051

Hollander, M. A., Gelman, S. A., & Star, J. (2002). Children's interpretation of generic noun phrases. *Developmental Psychology*, *38*(6), 883–894. https://doi.org/10.1037/0012-1649.38.6.883

Johnson, C. I., & Mayer, R. E. (2012). An eye movement analysis of the spatial contiguity effect in multimedia learning. *Journal of Experimental Psychology: Applied*, *18*(2), 178–191. https://doi.org/10.1037/a0026923

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. (2009). Transfer of mathematical knowledge: The portability of generic instantiations. *Child Development Perspectives*, *3*(3), 151–155. https://doi.org/10.1111/j.1750-8606.2009.00096.x

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, J. A., Sloutsky, V. M., & Heckler, A. F. (2013). The cost of concreteness: The effect of nonessential information on analogical transfer. *Journal of Experimental Psychology: Applied*, *19*(1), 14–29. https://doi.org/10.1037/a0031931

Keehner, M., Hegarty, M., Cohen, C., Khooshabeh, P., & Montello, D. (2008). Spatial reasoning with external visualizations: What matters is what you see, not whether you interact. *Cognitive Science: A Multidisciplinary Journal*, *32*(7), 1099–1132. https://doi.org/10.1080/03640210801898177

Kochari, A., Van Rooij, R., & Schulz, K. (2020). Generics and alternatives. *Frontiers in Psychology*, *11*, 1274. https://doi.org/10.3389/fpsyg.2020.01274

Leshin, R. A., Leslie, S.-J., & Rhodes, M. (2021). Does it matter how we speak about social kinds? A large, preregistered, online experimental study of how language shapes the development of essentialist beliefs. *Child Development*, *92*(4), e531–e547. https://doi.org/10.1111/cdev.13527

Leslie, S.-J. (2007). Generics and the structure of the mind. *Philosophical Perspectives*, *21*, 375–403.

Lopez, A., Gelman, S. A., Gutheil, G., & Smith, E. E. (1992). The development of category-based induction. *Child Development*, *63*(5), 1070. https://doi.org/10.2307/1131519

Mathiaparanam, O. N., Donovan, A., Menendez, D., Jones, C., Yoo, S. H., Alibali, M. W., Kalish, C. W., & Rosengren, K. (2022). Perceptual features in visual representations: A content analysis of inheritance diagrams. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *44*(44). https://escholarship.org/uc/item/9w66p0nz

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.

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

Menendez, D., Rosengren, K. S., & Alibali, M. W. (2022). Detailed bugs or bugging details? The influence of perceptual richness across elementary school years. *Journal of Experimental Child Psychology*, *213*, 105269. https://doi.org/10.1016/j.jecp.2021.105269

Moreno, R. (2007). Optimising learning from animations by minimising cognitive load: Cognitive and affective consequences of signalling and segmentation methods. *Applied Cognitive Psychology*, *21*(6), 765–781. https://doi.org/10.1002/acp.1348

Muradoglu, M., Marchak, K. A., Gelman, S. A., & Cimpian, A. (2022). Formal explanations shape children's representations of animal kinds and social groups. *Developmental Psychology*, *58*, 2322–2335. https://doi.org/10.1037/dev0001447

Noyes, A., & Keil, F. C. (2019). Generics designate kinds but not always essences. *Proceedings of the National Academy of Sciences*, *116*(41), 20354–20359. https://doi.org/10.1073/pnas.1900105116

Osherson, D. N., Smith, E. E., Wilkie, O., López, A., & Shafir, E. (1990). Category-based induction. *Psychological Review*, *97*(2), 185–200. https://doi.org/10.1037/0033-295X.97.2.185

Petersen, L. A., McNeil, N. M., Tollaksen, A. K., Boehm, A. G., Hall, C. J., Carrazza, C., Carrazza, C., Devlin, B. L., & Devlin, B. L. (2014). Counting practice with pictures, but not objects. *Proceedings of the Cognitive Science Society*, 7.

Rhodes, M., Rizzo, M. T., Foster-Hanson, E., Moty, K., Leshin, R. A., Wang, M., Benitez, J., & Ocampo, J. D. (2020). Advancing developmental science via unmoderated remote research with children. *Journal of Cognition and Development*, *21*(4), 477–493. https://doi.org/10.1080/15248372.2020.1797751

Schnotz, W. (2014). Integrated model of text and picture comprehension. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (2nd ed., pp. 72–103). Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.006

Schüler, A. (2017). Investigating gaze behavior during processing of inconsistent text-picture information: Evidence for text-picture integration. *Learning and Instruction*, *49*, 218–231. https://doi.org/10.1016/j.learninstruc.2017.03.001

Schüler, A., Arndt, J., & Scheiter, K. (2015). Processing multimedia material: Does integration of text and pictures result in a single or two interconnected mental representations? *Learning and Instruction*, *35*, 62–72. https://doi.org/10.1016/j.learninstruc.2014.09.005

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. (2022). Realistic visualizations can aid transfer performance: Do distinctive shapes and descriptive labels contribute towards learning? *Journal of Computer Assisted Learning*, *38*(3), 681–691. https://doi.org/10.1111/jcal.12640

Skulmowski, A., Nebel, S., Remmele, M., & Rey, G. D. (2021). Is a preference for realism really naive after all? A cognitive model of learning with realistic visualizations. *Educational Psychology Review*. https://doi.org/10.1007/s10648-021-09638-1

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*. https://doi.org/10.1002/hbe2.190

Sloutsky, V. M., & Fisher, A. V. (2004). When development and learning decrease memory: Evidence against category-based induction in children. *Psychological Science*, *15*(8), 553–558. https://doi.org/10.1111/j.0956-7976.2004.00718.x

Sloutsky, V. M., Lo, Y.-F., Fisher, A. V., Sloutsky, V. M., Lo, Y. F., & Fisher, A. V. (2001). How much does a shared name make things similar? Linguistic labels, similarity, and the development of inductive inference. *Child Development*, *72*(6), 1695–1709. https://doi.org/10.1111/1467-8624.00373

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

Stull, A. T., Hegarty, M., Dixon, B., & Stieff, M. (2012). Representational translation with concrete models in organic chemistry. *Cognition and Instruction*, *30*(4), 404–434. https://doi.org/10.1080/07370008.2012.719956

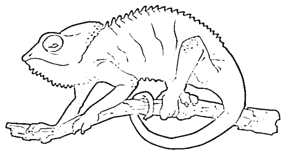
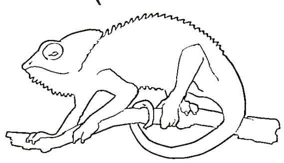
Tasimi, A., Gelman, S. A., Cimpian, A., & Knobe, J. (2017). Differences in the evaluation of generic statements about human and non-human categories. *Cognitive Science*, *41*(7), 1934–1957. https://doi.org/10.1111/cogs.12440

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

Viechtbauer, W. (2010). Conducting meta-analyses in *R* with the **metafor** package. *Journal of Statistical Software*, *36*(3). https://doi.org/10.18637/jss.v036.i03

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.

Picture	Text
Bland (Studies 2-5): 	Generic: Chameleons have sticky spit.
Bland (Studies 6-8): 	Exemplified generic: Chameleons, like the one shown above, have sticky spit.
Rich (Studies 2-8): 	Non-generic: This chameleon has sticky spit.

Sample Item

Chameleons are reptiles.
Chameleons have sticky spit.

How many other chameleons do you think have sticky spit?

None	A few	Some	Most	All
O	O	O	O	O

How many other reptiles do you think have sticky spit?

None	A few	Some	Most	All
O	O	O	O	O

Figure 1. Sample stimuli (bottom panel) and visualization and text examples (top panel) for the chameleon item.

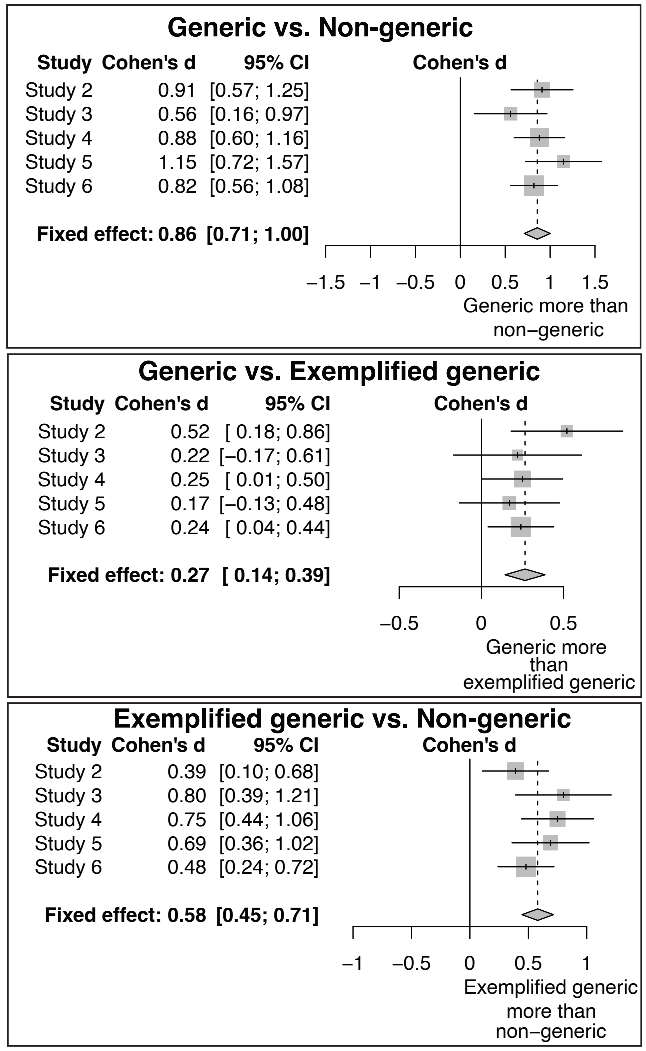


Figure 2. Forest plot for the internal meta-analyses of the effects of text type on generalization. The figure shows the study effect size estimates (Cohen's *d*) with their 95% confidence intervals. The size of the square on the forest plot shows the weight that each study was assigned. The diamond shows the meta-analytic estimate and the dotted line shows the mean of this estimate. The solid line shows 0 (no effect). The top panel shows the results for the generic vs. non-generic contrast. The middle panel shows the results for the generic vs. exemplified generic contrast. The bottom panel shows the results for the exemplified generic vs. non-generic contrast.

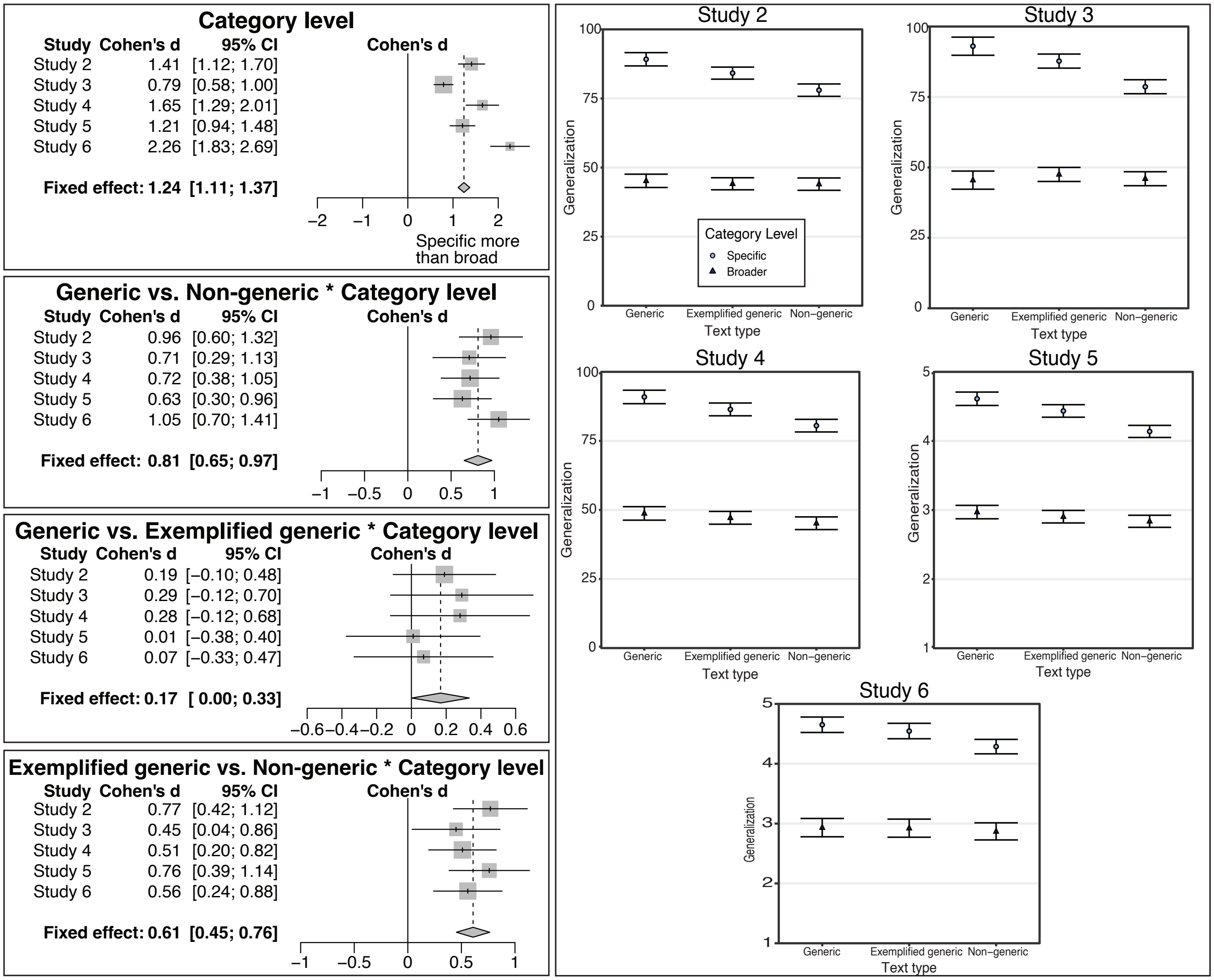


Figure 3. The left panel shows the forest plot for the internal meta-analyses of category level and the interactions between text type and category level on generalization. The figure shows the study effect size estimates (Cohen's *d*) with their 95% confidence intervals. The size of the square on the forest plot shows the weight that each study was assigned. The diamond shows the meta-analytic estimate, and the dotted line shows the mean of this estimate. The solid line shows 0 (no effect). In the left panel, from top to bottom the panels show the results for: category level, category level by generic vs. non-generic contrast, category level by generic vs. exemplified

generic contrast, and category level by exemplified generic vs. non-generic contrast. The right panel shows the results of the individual studies showing the interaction between text type (x-axis) and category level (color and shape). The error bars show the within-subject standard error of the point estimate.

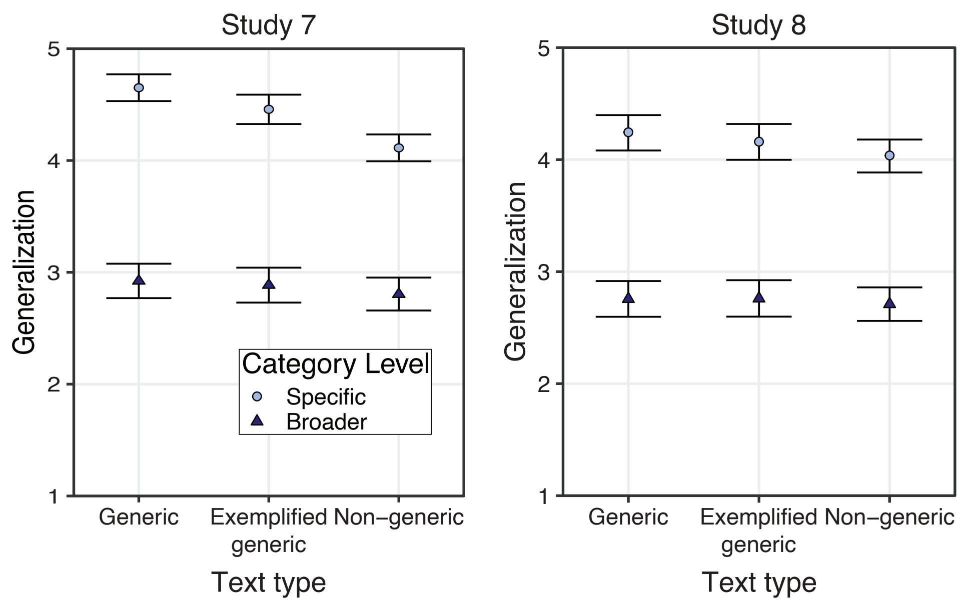


Figure 4. Model predictions of participants' responses to "How many other [animals] do you think [statement]?" broken down by text type (x-axis), and category level (shape). The left panel shows the results for Study 7 and the right panel the results for Study 8. The error bars show the upper and lower ends of the highest density interval.

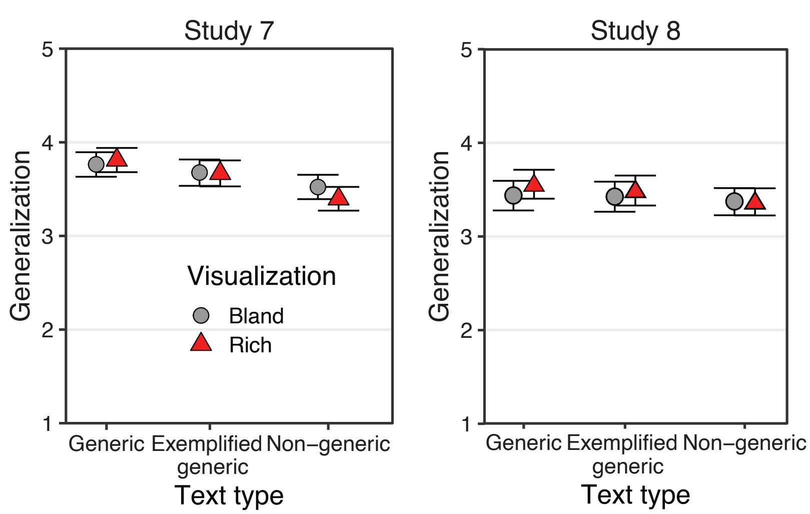


Figure 5. Model predictions of participants' responses to "How many other [animals] do you think [statement]?" broken down by text type (x-axis), and visualization type (shape and color). The left panel shows the results for Study 7 and the right panel the results for Study 8. The error bars show the upper and lower ends of the highest density interval.

Table 1

Model comparisons for Study 7

Model	Δelpd	SE
Baseline: fixed intercept + by-subject random intercepts + by-subject random slopes for text type, category level and their interaction + by-item random intercepts + by-item random slopes for text type, visualization, category level and all their interactions	-9.2	4.3
+ Category level	-9.5	4.8
+ Text type	-11.6	5.0
+ Visualization	-12.1	5.0
+ Text type x Visualization	-11.7	4.6
+ Text type x Category level	**0.0**	**0.0**
+ Visualization x Category level	-0.2	0.4
+ Text type x Visualization x Category level	-2.2	0.6

Note. Rows shows the different models and the order reflects the order in which they were fit (each model adding predictors to the previous one). Columns show the difference in the expected log-predicted likelihood between the model and the best fitting model(Δelpd) and the standard error (SE). The best fitting model has the highest elpd and is shown in bold.

Table 2

Model comparisons for Study 8

Model	Δelpd	SE
Baseline: Grade + by-subject random intercepts + by-subject random slopes for text type, category level and their interactions + by-item random intercepts + by-item random slopes for text type, visualization type, category level, grade and their interactions	-9.8	4.7
+ Text type	-3.9	3.6
+ Visualization type	-4.3	3.6
+ Category level	-1.5	3.0
+ Text type x Visualization type	-1.5	2.5
+ Text type x Category level	**0.0**	**0.0**
+ Text type x Grade	-0.2	1.6
+ Text type x Visualization type x Category level	-2.3	1.7
+ Text type x Visualization type x Category level x Grade	-0.9	3.8

Note. Rows shows the different models and the order reflects the order in which they were fit (each model adding predictors to the previous one). Columns show the difference in the expected log-predicted likelihood between the model and the best fitting model(Δelpd) and the standard error (SE). The best fitting model has the highest elpd and is shown in bold.