Published in final edited form as:

Dev Psychol. 2012 May ; 48(3): 890–896. doi:10.1037/a0026075.

Effects of categorical labels on similarity judgments: A critical analysis of similarity-based approaches

Nicholaus S. Noles and Susan A. Gelman

University of Michigan

Abstract

The goal of the present study was to evaluate the claim that category labels affect children's judgments of visual similarity. We presented preschool children with discriminable and identical sets of animal pictures and asked them to make perceptual judgments in the presence or absence of labels. Our findings indicate that children who are asked to make perceptual judgments about identical items judge discriminable items less accurately when making subsequent similarity judgments. Thus, labels do not generally affect children's perceptual similarity judgments; rather, children's reliance on labels to make similarity judgments appears to be attributable to flaws in the methodological approaches used in prior studies. These results have implications for the role of perceptual and conceptual information in children's categorization and induction.

Categorization and induction are powerful cognitive processes that help children interpret novel situations and infer information that they have not been explicitly taught. For example, if a child learns that the four-legged animal that lives in her house is a "dog," then she might effectively use this knowledge to identify other dogs, or to apply what she has learned about Fido to other animals that share the same features or the same label (for a comprehensive review, see Murphy, 2002).

Researchers have employed a number of approaches in exploring the cognitive architecture that underlies children's categorization and inductive inferences. Similarity-based and associative learning approaches focus on children's abilities to detect and correlate perceptual (e.g., Sloutsky & Napolitano, 2003) and statistical (e.g., Rakison, 2004) information in their environment. In contrast, theory-based approaches highlight the salience of children's knowledge and intuitive theories regarding intentions (e.g., Wellman & Phillips, 2001), causality (e.g., Gopnik & Sobel, 2000; Opfer & Bulloch, 2007), and other non-obvious properties (e.g., Gelman, 2003). Although it has been suggested that the competition between these two viewpoints is founded on a series of false dichotomies between perceptual and conceptual factors (e.g., Waxman & Gelman, 2009), and that both knowledge-based and similarity-based factors influence children's early word and concept learning, theorists differ in the extent to which they emphasize one or the other position.

Similarity-based approaches are elegant in their simplicity; they rely on only basic perceptual and associative systems to explain the acquisition and application of knowledge. However, similarity-based approaches face a challenge in characterizing certain aspects of concepts, such as feature centrality (some features are more important for categorization than others, and these features are often non-obvious, see Gelman & Wellman, 1991) and context sensitivity (the importance of specific features changes across categories; e.g., curvature is a salient feature of boomerangs but not bananas, see Medin & Shoben, 1988),

which knowledge-based theorists attribute to children's understanding of deep conceptual principles. Similarity-based approaches also face a challenge when asked to explain the phenomenon that children seem to overlook perceptual similarity and rely instead on category membership when making inductions about entities (e.g., children use category membership to infer that two animals share properties; see Gelman & Markman, 1986).

A number of similarity-based models have been proposed to account for such findings, focusing on attentional biases (e.g., Jones & Smith, 1993; Rakison, 2000) or correlations between specific features (McClelland & Rogers, 2003; Rakison & Hahn, 2004). However, these models largely conceptualize language as a factor that orients or shifts attention to visual stimuli, and they require ongoing adjustments in order to account for children's inductive inferences and similarity judgments over the course of development. One model, however, treats language differently. According to the similarity-based view proposed by Sloutsky and Fisher (2004), words function as perceptual features in an undifferentiated similarity judgment, to be combined with other perceptual features (including appearances; Sloutsky, 2003). We refer to this as the words-as-features view, and it stands out because it accounts for influential labeling effects and provides a principled explanation for changes in children's behavior over time.

In contrast, the knowledge-based view maintains that words refer to concepts, and it is those concepts (not the uttered words per se) that mediate children's judgments (see Gelman & Coley, 1990; Gelman & Waxman, 2007; Jaswal, 2004; Keates & Graham, 2008). For example, Jaswal and Markman (2007) demonstrated that children used labels to make inferences about entities, even when those labels were not congruent with the perceptual features of the entity (e.g., a cat-like animal identified as a "dog" was inferred to eat bones instead of drinking milk). Children make category-based inferences even when labels are synonymous but not identical (e.g., puppy, dog; Gelman & Markman, 1986). Children appear to believe unexpected labels rather than merely being compliant, as they pass the label along to a new person who hadn't been present during the labeling event (Jaswal, Lima, & Small, 2009). Furthermore, children do not use labels indiscriminately, but use speaker certainty to determine whether the label should be used as the basis of induction (Jaswal, 2004). These data collectively argue in support of the knowledge-based view.

Thus, a key point of contention is the relationship between shared labels and judgments of similarity. According to the words-as-features view, any shared label should lead to judgments of increased perceptual similarity. In contrast, according to the knowledge-based view, similarity is not global and undifferentiated, but rather weighted by the dimensions of task relevance. On this view, labels contribute to conceptual similarity but not visual similarity.

Prior research found support for the words-as-features view. In these experiments, participants saw a series of picture triads, each consisting of a target item and two test options (A and B) (Sloutsky & Fisher, 2004). Children were asked to tell the experimenter whether test A or test B looked more like the target. Test A was identical to the target item; the similarity between test B and the target item varied between trials. Results indicated that, when labels were provided, children's judgments of visual similarity were influenced by the labels (i.e., B was judged to be relatively more similar to the target, if the target and B shared a label).

However, past tests of whether labels contribute to children's visual similarity judgments had two primary limitations: (a) test trials included items for which different instances were perceptually indistinguishable, thereby encouraging children to use whatever differentiating feature was at hand (in many cases, labels were the only distinguishing features), and (b)

children were not trained to ensure that they understood that their judgment should focus exclusively on visual perceptual cues.

Inclusion of perceptually indistinguishable test items

In prior work, four levels of similarity between test B and the target item were presented to participants. The most similar triads featured three identical items (e.g., three identical pictures of the same dog); the most dissimilar triads consisted of two identical items and an obviously dissimilar item (e.g., a target dog paired with an identical dog and a morphed “cat-dog”; see Sloutsky & Fisher, 2004). However, although four different levels of similarity were included, data from a No-label condition indicate that, for three of the four levels, preschoolers responded close to chance levels. Although comparisons to chance are not provided in the paper, means of matching test B to the target item (i.e., the incorrect response) ranged from above 40% to just below 50%, suggesting that children had great difficulty determining which picture was more similar. Children consistently matched the target item to test A only when test B was at the lowest level of similarity, which was the case in only 25% of the test trials. Thus, the label was by far the more salient differentiating element in 75% of the test trials. This greater cue validity of labels versus pictures in the majority of trials may have influenced children’s judgments, encouraging them to use the label when it was available.

Lack of training

A second limitation of previous research was that participants received no training to ensure that the task instructions were clear. Specifically, prior to receiving test trials, participants received only two warm-up trials, and no feedback. Thus, children were not trained to ensure that they understood that their judgment should focus exclusively on *visual* cues. Obviously this is crucial to the goals of the research, because if children interpret the task as involving a judgment of all modalities (auditory as well as visual), then they should take into account the label. This issue was exacerbated in some studies, in which the wording implied global similarity (e.g., “more similar to”; Sloutsky & Lo, 1999). But even when the wording was more precise, children did not receive instructions to clarify the purpose or goals of the task. Since children received no feedback throughout the experimental session, they could have adopted a strategy of simply matching the labels, and the experimenter would have no opportunity to assess this.

The present study was designed to examine the claim that shared labels enhance the perceived visual similarity between items. We replicated the work of Sloutsky and Fisher (2004) with a few important alterations to their methodology. First, we limited the study to two sorts of triads: identical triads and discriminable triads. Second, we counterbalanced the order of these item types, in order to assess the effect (if any) of including visually indiscriminable triads. Third, we modified the instructions so that they contained minimal ambiguity and additional references to the visual nature of the matching task. Finally, we added a training task to our two basic conditions in order to ensure that children understood that they were being asked to make visual similarity judgments. A third condition, in which children received no training, was also created in order to determine if the training influenced label use. Our goal was to present children with three conditions that were constructed of the same visual and auditory elements used in previous research, while eliminating many of the confounding aspects found in the original design.

Method

Participants

We tested a total of 99 children and 35 adults. Twenty-six 4-year-olds (12 female and 14 male, $M = 4.63$, $SD = .62$) were tested in the No-label condition, 28 4-year-olds (10 female and 18 male, $M = 4.55$, $SD = .40$) were tested in the Label-plus-training condition, and 32 4-year-olds (15 female and 17 male, $M = 4.30$, $SD = .89$) were tested in the Label-minus-training condition. In order to make certain that we included data only from children who understood the task in our analyses, we excluded data provided by three children whose training accuracy was below 75%. Thirty-five adults and 10 children participated in pretesting of the stimulus materials (see below). Child participants were recruited from local day care centers in a small university town in the Midwestern United States, representing a wide variety of social and economic backgrounds, though most participants were White. Adults were tested in a laboratory setting and received course credit for participation; children were tested at their school and received a sticker for their efforts.

Materials

Pictures of geometric shapes, foods (e.g., pumpkins, doughnuts, apples, pears), and artifacts that were visually similar to different foods (e.g., a tire that resembles a chocolate doughnut and a red balloon that looks like an apple) were used as stimuli for the training tasks. The pictures of food and artifacts were used to create triads pitting visual similarity against category membership (e.g., a red whole apple shares perceptual features with a red balloon, but a whole red apple and a cut-up green apple are the same kind of food). Test items were created by Sloutsky and Fisher using pictures of real and morphed animals. These stimuli were used to create two different kinds of triads: discriminable and identical. Each triad consisted of a target item and two test items. The animal pictures were a subset of the images that were used by Sloutsky and Fisher in their original investigations, including triads with similarity ratios of 1 (three identical images) and 9 (two identical images and a morphed image). Our pictures differed from those employed by Sloutsky and Fisher in one respect. In prior work, some pictures were presented more than once across similarity ratios (e.g., a perceptual match in an indiscriminable trial might be used as a lure in a dissimilar triad, resulting in children selecting the item as a match in one trial and ignoring it as a distractor in another). In contrast, no pictures were presented in more than one triad in the present study. Critically, we did not change the discriminable triads; we only replaced repeated images in identical triads with other pictures from Sloutsky and Fisher's materials.

Pretesting

In order to select the training items, 20 similarity-versus-category triads were pretested with both children and adults, with the goal of selecting the best triads for the training portion of the study proper. We asked 20 adults to rate the perceptual similarity between the target item and each of the two test items of each practice triad. Of 20 initial triads, we selected the 14 triads with the highest similarity ratings between the target and perceptual match and the lowest similarity ratings between the target and category match. We then pretested these 14 sets with ten children ($M = 4.41$, $SD = .32$). On the pretest, children were asked to name each of the images. If they could not name an image, or if they did so incorrectly, they were asked to determine whether the image depicted something that was "food" or "not-food." Our intention was to make certain that children could easily make the categorical distinction between images depicted in these triads. The children were able to name only about half of the items ($M = 51\%$), but their food/not-food judgments were very accurate ($M = 88\%$). Two triads were eliminated from consideration because children found the pictures particularly difficult to name, and two triads were eliminated because of excessive similarity between the target and the 'dissimilar' test item (e.g., an egg and an egg timer shaped like an

egg were too alike). Of the remaining 10 item sets (see Table 1), two were used to model the correct answers and the remaining eight triads were used as the training stimuli (conflict triads) in the experimental sessions. Of these eight stimuli, four were “easy” to name (i.e., seven or more participants named two or more of the items in a given triad), and four were “hard” to name (i.e., fewer than seven participants could name at least two of the items in a given triad).

Procedure

In the Label-plus-training and No-label conditions, each experimental session involved four phases: 1) warm-up, 2) modeling, 3) training, and 4) testing. Each participant was presented with a short introduction to the task. The experimenter said, “We’re going to play a game today. In this game we’re going to look at some pictures. I’m going to show you three pictures and it’s going to be your job to pick which of the bottom pictures looks the most like the one on the top. Some of these are tricky, so make sure that you pay attention to the pictures. Ready?” The experimenter then began the warm-up phase, which consisted of four trials requiring children to match colors and shapes (e.g., two blue squares and a red circle). This part of the experiment was designed to familiarize children with the visual nature of their task. The experimenter began the task, saying, “Can you tell me which of these two [point to test items] looks the most like this one [target item]? Which one looks the same when you look at it with your eyes?” If children were hesitant to answer in any phase of the experiment, including training or test, they were reminded with, “Which one looks the same when you look at it with your eyes?” in order to emphasize the visual nature of the task. Both the warm-up and training phases were presented with feedback. If a child failed to select the visual match, the experimenter responded, saying, “Oops! I don’t think that’s right. I think that this one [indicates visual match] looks more like this one [target item].”

After completing the warm-up phase, the modeling phase began. In the modeling phase, the experimenter completed two example triads in order to demonstrate the desired response to the child. The experimenter introduced this phase, saying, “Good job! Now we’re going to do some that are a little trickier. I’ll do the first couple and then it will be your turn.” Then, for each of two modeling trials, the experimenter said, “When I look at these with my eyes, I think that this one [visual match] looks like this one [target].” The experimenter then indicated that it was the child’s turn to play the game and began the training phase.

In the training phase, the experimenter repeated the instructions used in the warm-up phase, emphasizing the importance of making judgments “with your eyes.” The training phase consisted of twelve trials. Eight conflict trials pitted visual similarity against category membership (e.g., a chocolate doughnut could be matched with either a tire [visual similarity] or a jelly doughnut with a bite taken out [category membership]), and four no-conflict trials pitted same-category visual similarity against a distracter (e.g., a chocolate-chip cookie could be matched with either another chocolate-chip cookie [same category, visual similarity] or a metal fly wheel [distracter]). We presented same-category visual matches in order to be certain that children didn’t apply non-perceptual heuristics to the training task. For example, without same-category matches, children might simply apply a strategy where they selected non-foods.

The procedures in the two conditions differed in the test phase. First, all children heard the following instructions: “Now, we’re going to play a game about animals from faraway places. I’m going to show you pictures of animals, and then ask you to point to them. Then, just like we did before, I’m going to ask you to tell me which ones look more alike when you look at them with your eyes.” In the No-label condition, the experimenter pointed to each of the images and said, “Look at this one.” After pointing to the pictures in a triad, the experimenter prompted the child to point to each picture and then said, “Look at them

carefully with your eyes. Does this one [target] look more like this one [test item] or more like this one [test item]?” Children in the Label-plus-training condition heard the same instructions, except that each picture was labeled (selected from the set of labels used by Sloutsky and Fisher; lolo, tippy, jaja, etc.). As in Sloutsky and Fisher, the target items shared labels with the non-visual matches within each triad. Thus, for the discriminable sets, the labels conflicted with visual similarity (i.e., picture that was identical to the target received a different label, whereas the picture that was different from the target received the same label). For the identical sets, the two test pictures received different labels, so that the target received the same label as one picture and a different label from another picture. Also, instead of pointing to the images, participants in the Label-plus-training condition were required to correctly name each image before making their similarity judgment. After the child correctly named all three images, the experimenter said, “Look at them carefully with your eyes. Does this [label 1, e.g., *lolo*] look more like this [label 2, e.g., *tippy*] or more like this [label 1, e.g., *lolo*]?” Participants received eight test trials. The Label-minus-training condition was identical to the test phase of the Label-plus-training condition.

Items presented in the training phase were randomized, with half of the participants in each condition receiving them in one order and the other half receiving them in the opposite order. The test trials were blocked so that children received one block of four identical triads and one block of four discriminable triads, in counterbalanced order. Triads in both training and test phases were constructed so that perceptual matches appeared equally on the left and right sides of triads, in order to prevent children from favoring specific test locations. All materials were presented in booklets.

Results and Discussion

During the training phase, children accurately matched perceptually similar images in both the No-label ($M = 94\%$) and Label-plus-training conditions ($M = 95\%$), demonstrating that they understood the nature of the visual matching task. When children did make errors on the training task (which was quite rare), it did not appear to be related to their ability to name the items: Children made errors on 8% of triads that were easily named at pretest, and they made errors on 5% of triads that were difficult to name on pretest.

For the test trials, children's reliance on labels when making similarity judgments was evaluated using a 3 (Condition) $\times$ 2 (Order) $\times$ 2 (Similarity) mixed analysis of variance with Condition and Order as a between subjects factor and Similarity as a repeated measure. The dependent measure was the proportion of trials on which children selected the same-label item, ranging from 0 (child always relied upon perceptual similarity) to 1 (child always selected test items that shared labels with target items). Because children in the No-label condition did not actually receive labels, their “label selection” was determined by the percentage of the time that they selected the item that shared the same label in the yoked Label-plus-training condition. Children's responses are depicted in Figure 2. This test revealed main effects for Condition ($F(2,80) = 6.61, p = .002, \eta^2 = .142$), Order ($F(1,80) = 8.76, p = .004, \eta^2 = .099$), and Similarity ($F(1,80) = 131.78, p < .001, \eta^2 = .622$), as well as a significant Order $\times$ Similarity interaction ($F(1,80) = 12.89, p = .001, \eta^2 = .139$).

As expected, children used labels more often when triads were identical than when they were discriminable (65% versus 26%, respectively). Also as expected, when items were identical, children responded at chance in the No-label condition but used labels when they were available. Importantly, however, children's selections were also influenced by Order. When identical items were presented first, children relied on labels on 53% of trials, compared to using labels on only 38% of trials when discriminable items were presented first. Moreover, the Order $\times$ Similarity interaction reveals that children were most accurate

on the discriminable triads when they were presented first. In contrast, when identical triads were presented first, children's accuracy on the subsequent discriminable triads declined. Thus, being asked to make perceptual judgments on a series of identical triads (i.e., an impossible task) disrupted performance when the triads were discriminable.

Finally, the Label-minus-training and Label-plus-training conditions differed significantly from the No-label condition ($p < .05$), but they did not significantly differ from one another ($p = .77$). Thus, training did not make the children more or less likely to employ labels (i.e., performance was equivalent in the Label-plus-training and Label-minus-training conditions), implying that training was not required for children to make appropriate visual choices in the presence of a label.

For our purposes, the key finding is that children were much more accurate at making perceptual judgments in one task order than the other. As noted earlier, when discriminable triads appeared first, children performed more accurately than when identical triads appeared first. Conversely, when identical triads appeared first, children were much more likely to respond incorrectly. Why were children so much more likely to use the label in the "disfavored" order (when identical triads appeared first)? We suggest that when children received a series of identical triads, coupled with instructions to judge the items based on perceptual similarity, they were essentially given an impossible task. They therefore adopted an alternative strategy: either guessing (in the case of the No-Label condition) or relying on the label as the only available cue (in the case of the Label-plus-training conditions). This strategy then tended to be carried over into the second block of trials. Importantly, these data undermine the claim that labels contribute to the perceptual similarity of items (Sloutsky & Fisher, 2004), as labels influence perceptual similarity only under the adverse order of conditions.

The purpose of the present study was to examine whether category labels affect the perceived visual similarity of pictures, for preschool children. On the surface, our data replicate the results reported in Sloutsky and Fisher (2004), indicating that labels influence similarity judgments. However, when these data were broken down by presentation order (e.g., see the Label-plus-training condition data depicted in Figure 3), the pattern that emerges is inconsistent with the conclusions offered in prior studies by Sloutsky and colleagues. Our manipulation of block order revealed a general effect, wherein judging the similarity of identical items resulted in poorer similarity judgments of discriminable items. Although there was no significant 3-way interaction involving Condition, we were specifically interested in whether children in the Label-plus-training condition made significant use of the label on discriminable items, when discriminable items were presented first. Recall that the key prediction of the Sloutsky and Fisher argument was that use of the label will influence children's perceptual judgments. We therefore wished to know whether this effect holds up when the order did not encourage alternative strategies, as discussed above. We tested this by comparing children's responses in the two label conditions (Label-plus-training, Label-minus-training) to their responses in the No-label condition, for the discriminable items when receiving the discriminable items first. If the label had an effect on their judgments, then children should have used the label more frequently in the two label conditions than in the No-label condition. Importantly, however, Bonferroni analyses revealed that the presence of a label did not significantly influence children's similarity judgments of discriminable triads, when discriminable triads were presented first (6% in the No-label condition; 16% in the Label-plus-training condition; 17% in the Label-minus-training condition, $p > .12$).

Taken together, this collection of findings calls into question models of induction that assume that effects of category labels are due to perceived similarity. Children's similarity

judgments are affected by labels only when a block of identical triads is presented first. When discriminable triads are presented first, children accurately rely on perceptual similarity, and are uninfluenced by the competing labels. The primary advantage of the words-as-features view is that it models children's behaviors, even as they change over the course of development. Instead, our data indicate that the methodologies used to argue for this position (e.g., Sloutsky & Fisher, 2004) elicited certain response patterns, rather than measuring them. Minimally, our data suggest that perceptual and associative processes alone are insufficient to describe the rich set of behaviors exhibited by children. More generally, the present findings are consistent with other research showing that labels link to conceptual knowledge and do not reduce to perceptual similarity alone (e.g., Jaswal & Markman, 2007; Opfer & Bulloch, 2007; Preissler & Carey, 2004). We suggest that children make use of multiple cues when learning words and drawing inductive inferences, including perceptual similarity, associative learning, knowledge-based causal theories, and social information (see also Waxman & Gelman, 2009).

Acknowledgments

For assistance and helpful conversation and/or comments on earlier drafts, we thank three anonymous reviewers, Judith Danovitch, Margaret Evans, Liza Ware, Henry Wellman, and the members of the Language Lab at the University of Michigan. We would also like to thank Vladimir Sloutsky for providing copies of his visual stimuli. Thank you to the children, teachers, administrators, and staff of Annie's Children's Center, Child time Children's Center, The Discovery Center, Little Folks Corner, Perry Nursery School, Play and Learn Children's Place, and Stony Creek Preschool Too. This work was supported by NICHD grant HD-36043 awarded to Gelman. We thank Nick Bauer-Levy, Gemma Bransdorfer, Liwen Chen, Taryn Konevich, Jessica McKeever, Meghan McMahon, Elana Mendelowitz, Miyuki Nishimura, Alex Picurko, Kathryn Pytiak, Sarah Stilwell, and Alex Was, for research assistance.

References

- Gelman, SA. The essential child: Origins of essentialism in everyday thought. New York: Oxford University Press; 2003.
- Gelman SA, Coley JD. The importance of knowing a dodo is a bird: Categories and inferences in 2-year-old children. *Developmental Psychology*. 1990; 26:796–804.10.1037//0012-1649.26.5.796
- Gelman SA, Markman EM. Categories and induction in young children. *Cognition*. 1986; 23:183–209.10.1016/0010-0277(86)90034-X [PubMed: 3791915]
- Gelman SA, Waxman SR. Looking beyond looks: Comments on Sloutsky, Koos, & Fisher. *Psychological Science*. 2007; 18:554–555.10.1111/j.1467-9280.2007.01937.x [PubMed: 17576270]
- Gelman SA, Wellman HM. Insides and essences: Early understanding of the non-obvious. *Cognition*. 1991; 38:213–244.10.1016/0010-0277(91)90007-Q [PubMed: 2060270]
- Gopnik A, Sobel DM. Detectingblickets: How young children use information about novel causal powers in categorization and induction. *Child Development*. 2000; 71:1205–1222.10.1111/1467-8624.00224 [PubMed: 11108092]
- Jaswal VK. Don't believe everything you hear: Preschoolers' sensitivity to speaker intent in category induction. *Child Development*. 2004; 75:1871–1885.10.1111/j.1467-8624.2004.00822.x [PubMed: 15566385]
- Jaswal VK, Lima OK, Small JE. Compliance, conversion, and category induction. *Journal of Experimental Child Psychology*. 2009; 102:182–195.10.1016/j.jecp.2008.04.006 [PubMed: 18556016]
- Jaswal VK, Markman EM. Looks aren't everything: 24-month-olds' willingness to accept unexpected labels. *Journal of Cognition and Development*. 2007; 8:94–111.10.1080/15248370709336995
- Jones SS, Smith LB. The place of perception in children's concepts. *Cognitive Development*. 1993; 8:113–139.10.1016/0885-2014(93)90008-S
- Keates J, Graham SA. Category markers or Attributes: Why do labels guide infants' inductive inferences. *Psychological Science*. 2008; 19:1287–1293.10.1111/j.1467-9280.2008.02237.x [PubMed: 19121139]

- McClelland J, Rogers TT. The parallel distributed processing approach to semantic cognition. *Nature Reviews Neuroscience*. 2003; 4:1–13.10.1038/nrn1076
- Medin DL, Shoben EJ. Context and structure in conceptual combination. *Cognitive Psychology*. 1988; 20:158–190.10.1016/0010-0285(88)90018-7 [PubMed: 3365938]
- Murphy, GL. *The big book of concepts*. Cambridge: MIT Press; 2002.
- Opfer JE, Bulloch MJ. Causal relations drive young children's induction, naming, and categorization. *Cognition*. 2007; 105:206–217.10.1016/j.cognition.2006.08.006 [PubMed: 17045580]
- Preissler MA, Carey S. Do both pictures and words function as symbols for 18- and 24-month-old children? *Journal of Cognition and Development*. 2004; 5:185–212.10.1207/s15327647jcd0502_2
- Rakison DH. When a rose is just a rose: The illusion of taxonomies in infants' categorization. *Infancy*. 2000; 1:77–90.10.1207/S15327078IN0101_07
- Rakison DH. Infants' sensitivity to correlations between static and dynamic features in a category context. *Journal of Experimental Child Psychology*. 2004; 89:1–30.10.1016/j.jecp.2004.06.001 [PubMed: 15336916]
- Rakison, DH.; Hahn, E. The mechanisms of early categorization and induction: Smart or dumb infants?. In: Kail, R., editor. *Advances in child development and behavior*. New York: Academic Press; 2004. p. 281–322.
- Sloutsky VM. The role of similarity in the development of categorization. *Trends in Cognitive Sciences*. 2003; 7:246–251.10.1016/S1364-6613(03)00109-8 [PubMed: 12804690]
- Sloutsky VM, Fisher AV. Induction and categorization in young children: A similarity-based model. *Journal of Experimental Psychology: General*. 2004; 133:166–188.10.1037/0096-3445.133.2.166 [PubMed: 15149249]
- Sloutsky VM, Lo Y. How much does a shared label make things similar? Part 1. Linguistic labels and the development of similarity judgment. *Developmental Psychology*. 1999; 35:1478–1492.10.1037/0012-1649.35.6.1478 [PubMed: 10563736]
- Sloutsky VM, Napolitano AC. Is a picture worth a thousand words? Preference for auditory modality in young children. *Child Development*. 2003; 74:822–833.10.1111/1467-8624.00570 [PubMed: 12795392]
- Waxman SR, Gelman SA. Early word-learning entails reference, not merely associations. *Trends in Cognitive Sciences*. 2009; 13:258–263.10.1016/j.tics.2009.03.006 [PubMed: 19447670]
- Wellman, HM.; Phillips, AT. Developing intentional understandings. In: Malle, B.; Moses, L.; Baldwin, D., editors. *Intentionality: A key to human understanding*. Cambridge: MIT Press; 2001. p. 125–148.

Figure 1. Test Item Similarity Ratings

Test triads, originally created by Sloutsky and Fisher (2004), with adult similarity ratings. 1= “look very different” and 7 = “look very similar”. Test items in identical triads were judged to be significantly more similar to target items than test items in discriminable triads, $t(14) = 17.18, p < .001$.

Figure 2. Label Selection

Mean % of trials on which children selected the item with the same label as the target, when making similarity judgments

Figure 3. Order Effects

Mean % of trials on which children selected the item with the same label as the target across all conditions, as a function of block order.

Table 1

Training Triads

Phase	Target	Perceptual Match	Lure
Warm-up	Blue square	Blue square	Red circle
	Blue square	Blue square	Blue circle
	Blue square	Blue circle	Red triangle
	Blue square	Green square	Red circle
Modeling	Purple cabbage	Purple ornament	Bowl of salad
	Chocolate cookie	Black Frisbee	Heart cookie
Training-conflict triads	Baguette	Baseball bat	Slice of bread
	Chocolate doughnut	Tire	Jelly doughnut
	Whole red apple	Red balloon	Sliced green apple
	Whole cantaloupe	Ball of twine	Sliced cantaloupe
	Pumpkin	Basketball	Sliced zucchini
	Whole cheese pizza	Clock	Veggie pizzaslice
	Potato	Football	French fries
	Raspberry	Textured thimble	Blueberry
Training -no-conflict triads	Tomato	Tomato	Fire hydrant
	Pear	Pear	Baseball
	Cookie	Cookie	Metal fly wheel
	Apple	Apple	Tennis ball

As with the practice triads, we asked adults ($n=15$) to rate the similarity between the target and test items used in this study on a 7-point scale. The purpose of this test was to evaluate the degree of similarity in each discriminable triad, relative to the identical triads (see Figure 1).