

Saliency versus prior knowledge - how do children learn rules?

Samuel Rivera (rivera.162@osu.edu)

Department of Psychology, 1835 Neil Avenue
Columbus, OH 43210 USA

Vladimir M. Sloutsky (sloutsky.1@osu.edu)

Department of Psychology, 1835 Neil Avenue
Columbus, OH 43210 USA

Abstract

Categories are essential for thinking, learning, and communicating. Research has shown that young children and adults treat categories very differently, with young children favoring whole objects while adults focus on the key information in most cases. If so, then how can young children learn categories requiring focused attention to key features? Studies have shown that drawing attention to rules had facilitative effects. We sought to identify whether the effect was driven by instruction about rules or by stimulus-driven factors. Our results suggest that even with instruction, 4-year-olds were not able to attend to key information. Simply making important information more salient, however, allowed them to learn the category and *transfer* to situations when the key feature was no longer salient.

Keywords: category learning; attention; cognitive development

Introduction

Our world is made up of categories and concepts - groups of objects or ideas that share some equivalence. They allow us to generalize, communicate, and make decisions by abstracting away the unnecessary variation in the world. Decades of research have demonstrated that young children have difficulty attending selectively to "category-relevant" information and using this information when learning categories (Smith & Kemler, 1977). Children are thought to be *holistic* learners, with evidence showing a distributed pattern of attention in comparison to the focused attention profile of adults (Best, Yim, & Sloutsky, 2013). This means that children categorize objects based on their overall appearance while adults emphasize key features when categorizing. This phenomenon has been linked to the protracted development of selective attention and executive functioning in children (Smith, 1989; Sloutsky, 2010; Rabi & Minda, 2014). But if young children cannot focus on defining features, how and under what circumstances do they learn rule-defined categories? Answering this question is important for developing a theory of early categorization, but also has immediate practical relevance. Specifically, it informs efforts in teaching young children certain abstract concepts, such as fractions, where understanding a key idea or feature is prerequisite to learning.

Prior work has suggested that the solution to overcoming the holistic predisposition may depend on attention influences, with both instruction (Deng & Sloutsky, 2015) and feature saliency (Deng & Sloutsky, 2012; Rabi, Miles, & Minda, 2015) contributing to learning and generalization. Deng and Sloutsky demonstrated this point in their study of the role of

labels over development (Deng & Sloutsky, 2012). In their categorization and induction study, they found that when a very salient feature was in conflict with several other feature values, children would respond based on the salient feature value. Rabi and Minda (Rabi et al., 2015) similarly found that saliency had a significant effect on what type of information children used to categorize, and their success or failure in learning category rules. Their results suggested that children employed a variety of optimal rule-based, suboptimal rule-based, and similarity based strategies when children were learning categories defined by rules and family resemblance (R+FR). However, children were more successful at learning the rule when the rule-feature was salient. These results show that exogenous forces of the stimulus properties can facilitate category learning in young children.

In a more recent investigation of categorization and induction, Deng and Sloutsky (2005) found that 4-year-olds relied on overall similarity to categorize objects when they were learned by categorization or inference training. In inference training, the objective is to determine a missing feature value instead of the unknown category label as in classification training. When 4-year-olds were trained to categorize and repeatedly given instruction about the rule feature, however, they relied on the rule feature and categorized like 6-year-olds. This work implies a facilitative effect on top-down knowledge in selecting the key information for categorizing. Unfortunately, it is difficult to decouple the influence of instruction that requires effortful top-down control from selection history (Awh, Belopolsky, & Theeuwes, 2012) because instruction included specific exposure to features.

The short survey above demonstrates that instruction as well as feature saliency may have effects on young children's ability to learn categories. Our goal is to understand how and when those factors contribute to category learning. One possibility is that young children have difficulty identifying the important features when learning. Thus, the difficulty in focusing on key information during childhood could stem from poor learning strategies in general, which may be remedied with prior knowledge of key information. On the other hand, focusing on the key feature may not be possible early on due to limited selective attention. If that is the case, then an explicit manipulation of feature saliency would aid learning. A third possibility is that both instruction and saliency contribute to rule-based category learning to different degrees over development.

Thus, in this work we attempt to decouple the effect of both instruction, a form of prior knowledge, and stimulus properties (exogenous influences) on category learning in order to understand how those forces contribute to category and concept learning over development. To address these open questions, we investigated the differential influence of attentional cues on category learning and generalization. We predicted that exogenous cues would have a more potent facilitative effect on children’s category learning than endogenous cues due to limited executive functioning. Even with instruction, having to use effortful top-down control may not be as easy for young children. It was unclear whether endogenous instruction would help at all, or to a lesser degree. If young children can direct their attention to important features when learning categories, we would expect no difference between the exogenous versus endogenous cue. In the following we detail our experimental methodology and results. We close with a discussion.

Method

Participants

A total of 273 adults participated in the study for course credit through The Ohio State University research experience program. One additional adult participant who did not complete the study was excluded from the analysis. Adults either participated in the baseline (N = 74), endogenous (N=87), or exogenous (N=112) condition. A total of 87 four-year-olds ($M \pm SD = 4.52 \pm 0.27$ years) participated in the study. They completed either the baseline (N=24; $M \pm SD = 4.57 \pm 0.22$ years in range [4.04 : 4.87]), endogenous (N=33; $M \pm SD = 4.45 \pm 0.28$ years in range [3.84 : 4.90]), or exogenous condition (N=30; $M \pm SD = 4.57 \pm 0.28$ years in range [4.01 : 5.00]). Children were recruited through local daycares or preschools located in Columbus, Ohio, and public birth records. The majority of child and adult participants were Caucasian. Adults gave informed written consent prior to participation. Child participants gave verbal assent, and a caretaker gave written consent prior to the study.

Stimuli

Stimuli were artificial tree-like visual categories comprised of 6 spatially separated petals on branches extending from a central trunk, as shown in Figure 1. The stimuli spanned approximately 22 by 22 degrees of visual angle on the display. Neighboring petals were separated by approximately 8 degrees of visual angle from their centers, with about 2 degrees of empty space between the petals.

Categories were defined by the color and shape of the features according to a rule plus family resemblance (R+FR) structure. Specifically, the top right petal had a deterministic color and shape that perfectly determined the category. All other features varied probabilistically, such that all but one color and one shape was associated with a contrasting category. Thus, participants could learn the rule, or consider the overall appearance of features to correctly classify the objects

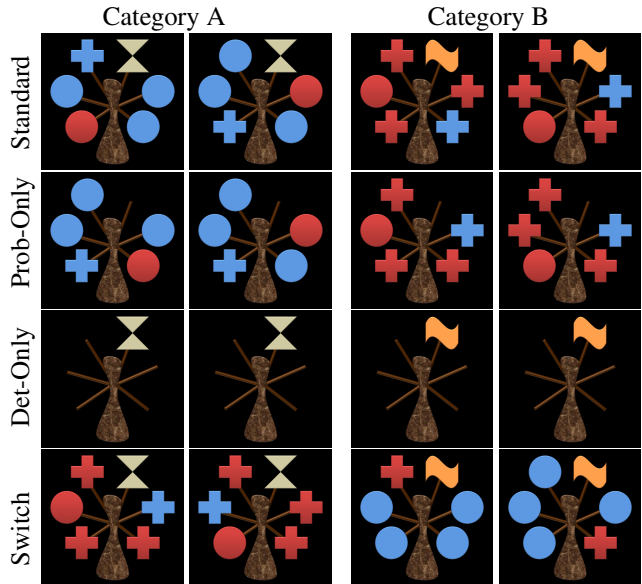


Figure 1: We show example stimuli used in the experiment. The first row shows standard stimuli used for the testing or training portion of the experiment. The bottom 3 rows show stimuli used only during testing. The Prob-Only images only show the probabilistic features or family resemblance features. The Det-Only condition displays only the rule feature. The switch items have the rule and family resemblance features in conflict.

Table 1: Example category structures used in the study. C1 and S1 correspond to color and shape at position one, respectively. Deterministic features are in bold.

	C1	S1	C2	S2	C3	S3	C4	S4	C5	S5	C6	S6
A	0	0	0	0	0	0	0	0	1	0	0	1
	0	0	0	0	0	0	1	0	0	1	0	0
	0	0	0	0	1	0	0	1	0	0	0	0
B	1	1	1	1	1	1	1	1	0	1	1	0
	1	1	1	1	1	1	0	1	1	0	1	1
	1	1	1	1	0	1	1	0	1	1	1	1

(Rosch & Mervis, 1975). Some example category structures are described symbolically in Table 1, where the feature value 0 corresponds to a value from category A, and a feature value 1 corresponds to category B.

Experiment

We ran a between subjects design where participants learned categories in either a baseline, endogenous cue, or exogenous cue condition. Testing was the same across conditions. The key manipulations are illustrated in Figure 2, with details explained below. Adults completed the task in a quiet room on campus at The Ohio State University. All instructions were displayed on the screen and read by participants at their own pace. Children completed the study one at a time with a trained experimenter reading the instructions and giving encouragement throughout.

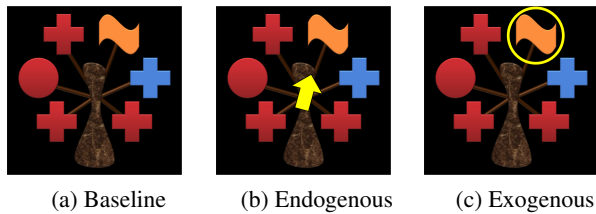


Figure 2: We show examples of stimuli during the training trials of the different experiment conditions. In the endogenous condition, the arrow remained during the entire trial but was not shown during test trials. Participants were told that the arrow points to an important feature that will help them categorize. In the exogenous condition, the yellow circle blinked during the training trials, but not during testing. Participants were told that some things may blink, but were not told that the feature is important.

Baseline In the baseline and all other conditions, participants learned to discriminate between two different categories. Participants were told that they would be playing a game where they would help creatures who live in either the desert or ponds to find their homes. Stimuli were shown centrally, with a small picture of a desert on the left, and a pond on the right. Adults responded at their own pace by pressing the left or right arrow which immediately prompted the feedback display. On the feedback screen the test exemplar was shown near the correct target (pond/desert), with either a central smiling or neutral face to indicate correct or wrong answers.

Participants first completed a series of 6 practice trials where they categorized frogs and camels. Each practice trial repeated until it was completed correctly. After practice, participants were told that they would fly away to a magical place with new types of creatures. They completed 30 training trials with feedback and an equal number of exemplars from each category. The left versus right position of the target was counterbalanced across participants. After the training trials, participants were asked to complete test trials without feedback. They had 8 standard test trials, 8 Switch trials, 8 probabilistic only (Prob-Only) trials, and 8 deterministic only (Det-Only) trials presented in a fixed order and without breaks between. Examples of stimuli from the different test trials are shown in Figure 1.

Child participants completed the same study, with a few age appropriate adjustments. Namely, children made their response by touching the pond or desert. In addition, only smiling faces were shown for correct answers and no face for incorrect answers. Corrective feedback was still given on every training trial in the form of the exemplar shown near the correct target on the feedback screen. The entire procedure took children about 10-12 minutes.

Exogenous The exogenous condition was like the baseline except for the key manipulation. Namely, during the instructions at the beginning of the training trials after practice, par-

ticipants were told that "Some things may blink but that is okay. Help them find their homes just like we did before." These instructions were given so that participants wouldn't be confused by the introduction of blinking after the initial practice trials. Then during the training trials, a yellow circle blinked around the deterministic feature. Nothing blinked on test trials.

Endogenous The endogenous condition was just like the baseline except for two key changes. During the post-practice instructions, adults were told that they "will see an arrow that will point to an important part of the creatures that will help you make your choice. Remember, the arrow points to something important!". Child instructions were more detailed. They were told that "There is a clue to this game! The clue will help you get all the smiley faces. You will see an arrow that will point to an important part of the creatures that will help you make your choice. Just like people who have different parts like arms and legs, the new creatures will have different parts. An arrow will point to a part that will help you know where it goes! Remember, the clue is that the arrow points to an important part that will help you know where the creature goes! Can you tell me the clue?" Experimenters repeated the instructions as necessary until children could mention the arrow clue and understood. About halfway through data collection, we added a memory check where we asked children if they remembered the clue. About 75% of polled children mentioned the arrow on the memory test. There was no effect of memory test result (pass/fail) on the training accuracy ($p=0.80$).

Results

We were interested in the differential effects of attention cues on learning and transfer to novel category exemplars. The different test trials probed different types of learning and response strategies. The standard test trials were an indicator of learning in general. Above chance (50%) accuracy on the standard test trials indicates successful category learning by either overall similarity or rule. The switch trials tell us how participants bias their decision. If the switch accuracy drops slightly from test but remains above chance, then participants may use both types of information but rely mostly on the deterministic feature. If the switch accuracy drops to below chance, then participants favor the probabilistic features in their decision. The Prob-Only trials probe whether participants had learned to use probabilistic features at all. If accuracy remained as high as on test, then we would infer that participants were using only this information. If it dropped slightly from test but remained above chance, then participants were using information from all features. If it dropped to chance on these trials, then we would infer that participants were only using the deterministic feature. On the other hand, the Det-Only trials tell us whether participants learned to use the rule feature. If accuracy dropped to chance, then we would infer that participants were using the probabilistic features to make their decision.

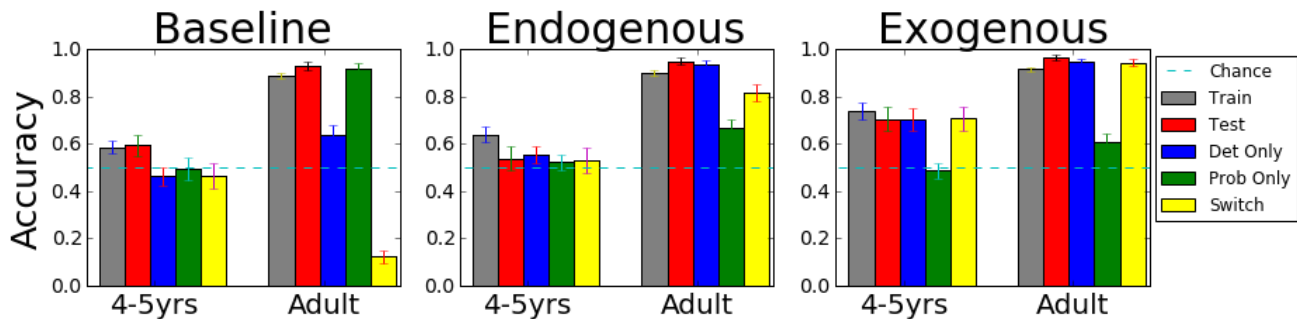


Figure 3: We plot mean accuracy across experiment conditions and age groups. Error bars denote standard error.

Our predictions were that adults would use rule features across all training conditions, with either cue providing equal support for learning that rule. That is because adults have been shown to optimize attention such that they only attend to the rule after learning (Shepard, Hovland, & Jenkins, 1961; Rehder & Hoffman, 2005; Hoffman & Rehder, 2010). We predicted that children, however, would receive more benefit from the exogenous versus endogenous cue, but would rely on the more holistic exemplar appearance when making their judgements. We expected their memory for the holistic appearance to be reduced in the exogenous condition versus the endogenous condition, from prior results showing the effects of very salient features on learning (Deng & Sloutsky, 2012).

The main analyses were concerned with the effects of the different training cues on accuracy for the different types of test trials, and changes in these effects over development. Therefore, we focused on participant accuracy within the different test trials. The accuracies are shown in Figure 3. We fit a 2x3x4 mixed ANOVA model with between subject factors of age (adult vs. child) and condition (baseline vs. endogenous vs. exogenous) and within subject factor of test type (Standard vs. Det-Only vs. Prob-Only vs. Switch) on accuracy. Analyses revealed a significant 3 way interaction, $F(4.66, 822.83) = 16.994, p < .001, \eta_p^2 = 0.088$ (Greenhouse-Geisser corrected for sphericity). We broke down by age group to further understand the effects.

For adults, we found a significant interaction between test type and condition, $F(4.26, 574.79) = 100.90, p < .001, \eta_p^2 = 0.43$ (Greenhouse-Geisser corrected for sphericity). We further broke down the interaction by condition. Standard test blocks across conditions were all above chance, indicating that adults successfully learned the categories in all conditions (baseline: $t(73) = 22.90, p < 0.001$; endogenous: $t(86) = 30.68, p < 0.001$; exogenous: $t(111) = 45.73, p < 0.001$; adjusted $\alpha = 0.002$ for 2x3x4 test type comparisons).

Pairwise comparisons within the adult baseline condition revealed that standard test accuracy and Prob-Only accuracy were not statistically different. Standard test accuracy was greater than Det-Only and Switch accuracy ($p < .001$), suggesting that adults were using probabilistic features in their categorization decisions. Surprisingly, we found that the re-

liance on the probabilistic features was overwhelming in the switch condition, with accuracy in Switch block being significantly lower than all other test blocks (pairwise comparison, $p < .001$). Contrary to our predictions, a one sample t-test revealed that Switch accuracy was significantly below chance level ($M=0.12$, one sample $t(73) = -12.93, p < 0.0001$). Thus, although participants could discriminate the categories above chance level with only the deterministic feature (Det-Only: $M=0.64$, one sample $t(73)=3.41, p=0.001$), in the presence of conflict they relied on the overall probabilistic appearance. We will revisit this novel and surprising finding in the Discussion.

Within the Endogenous condition, there was a different pattern. Standard test accuracy was not statistically different from the Det-Only, but was significantly better than the Prob-Only and Switch ($p < .001$). This suggests that adults in the endogenous condition relied on the deterministic feature overall, since they did as well in the Det-Only condition as in the Standard test block, but not as well when only probabilistic features were visible (Prob-Only). However, they did learn to rely on the probabilistic cue to a slight degree because accuracy was reduced from the standard test to switch trials - when deterministic and probabilistic features were in conflict. Unlike the baseline, however, the switch accuracy remained above chance indicating that the deterministic feature dominated in the decision.

Within the exogenous condition, the Standard test accuracy was greater than the Prob-Only condition ($p < .001$). However, Standard test accuracy was not significantly different from the Det-Only or Switch trials. This suggests that like the endogenous condition, adults relied on the deterministic feature. However, unlike the endogenous condition, there was no reduction in accuracy in the Switch, suggesting that probabilistic features were not used in the decision at all. Thus, the effect of salience was more pronounced than that of prior information in causing adults to learn and use the rule information.

We next investigated differences in the child age group. For children, we found a significant interaction between test type and condition, $F(5.76, 238.94) = 2.546, p = 0.023, \eta_p^2 = 0.058$ (Greenhouse-Geisser corrected for sphericity). We

further broke down the interaction by condition. Within the baseline and endogenous conditions, no test conditions were significantly different from each other ($ps > 0.18$). Pairwise comparisons confirmed that accuracy in the baseline and endogenous conditions were not significantly different from each other ($p > 0.99$). A one sample t-tests of standard test trials within the baseline and endogenous conditions revealed that accuracy was not significantly different from chance level. The baseline was approaching significance ($t(23) = 2.13, p = 0.044$), but not when adjusted for multiple comparisons (adjusted $\alpha = 0.002$). Thus, children did not successfully learn to discriminate the categories in the endogenous or baseline conditions.

Within the exogenous condition, however, a one sample t-test of accuracy in the standard test block confirmed that children successfully learned the categories ($t(29) = 4.058, p < 0.001$). Pairwise comparisons showed that only the Prob-Only accuracy was significantly less than all other testing conditions ($ps < .01$). One sample t-test revealed that accuracy in Prob-Only was not different from chance ($M = 0.4875, t(29) = -0.37, p = 0.71$). This suggests that children learned the category in the exogenous condition, and used the deterministic feature consistently in making their categorization judgements. When only probabilistic features were available, their accuracy dropped to chance performance.

These results are consistent with our prediction that children would do better learning the categories when salience draws their attention to the rule. As in studies mentioned above and similarly to adults in this study, the salience was overwhelming such that the probabilistic features were ignored in the decision. We should also emphasize here that the salient cue was not shown during the testing trials. Therefore, the information that was conveyed through a salient cue manipulation during training transferred to the non-cued test trials.

Discussion

We have presented a study to understand the conditions under which children can learn to use key information for categorizing objects. Decades of research has suggested that young children prefer holistic object information or overall similarity when categorizing objects, with a few more recent studies demonstrating facilitative effects of attentional manipulations on learning. We sought to disentangle the role of instruction and salience in these demonstrated effects. We make two important novel contributions.

First, we demonstrated a case where adults who learned the R+FR categories relied on the family resemblance features when categorizing. This demonstration is quite novel and surprising because decades of research has repeatedly shown that adults will use the rule features and ignore probabilistic features when learning categories with feedback. Furthermore, the R+FR category structure we employ is nearly identical to those used previously (See Table 1 vs. Deng & Sloutsky, 2012, Table 1). We suspect that the result stems from an im-

portant property that is obscured by the symbolic category structure representations. Namely, whereas the dimensions in other studies have distinct feature values, the family resemblance dimensions in our category structure denote highly redundant features. In other studies an exemplar with symbolic structure (000) corresponds to an object with distinct hands, feet, and body shapes that are all consistent with category A. In our study, however, a (000) corresponds to an exemplar having many blue features.

We suspect that adults in the baseline condition are sensitive to the redundancy, and learn to use *mostly blue circles* or *mostly red plus signs* as a type of higher order feature. There is evidence that adults do learn higher level features from primitives when learning categories and concepts (Schyns, Goldstone, & Thibaut, 1998). Furthermore, children are less consistent in their creation of these higher order features - possibly explaining their lesser reliance on this simple strategy. Interestingly, either attention cue made adults revert to the typical pattern of rule-based responding. Further research will be necessary to understand what strategy adults use in the baseline condition, and why they rely on the probabilistic features.

Our second main result is that instruction may not be sufficient for teaching young children to rely on key bits of information while simple salience is very powerful in that regard. We made a great effort in the endogenous condition to explain that the information necessary to the task would be in a location denoted by the arrow. We used analogies of body parts to convey the idea, and repeated the instructions as many times as necessary. However, children were not able to learn the categories given all those hints and guidance. In contrast, in the exogenous conditions we gave no hints or guidance. We only told participants that things would blink so that they would not be confused. We gave no instruction about the importance of blinking locations, but children learned that categories successfully and even transferred their knowledge to test trials where the blinking cue was not shown. The point that the knowledge transfers to the non-blinking case is critical because it implies that salience can be used to bootstrap concept learning for application in novel situations. Furthermore, detailed instruction may not be sufficient to overcome the limitations of limited executive function. Thus, our finding offers important insight into answering the question of how we can teaching young children concepts that require them to override their holistic predisposition.

Acknowledgments

This research was supported by NSF grant (BCS-1323963) to Vladimir Sloutsky.

References

- Awh, E., Belopolsky, A. V., & Theeuwes, J. (2012). Top-down versus bottom-up attentional control: A failed theoretical dichotomy. *Trends in Cognitive Sciences, 16*(8), 437–443. doi: 10.1016/j.tics.2012.06.010

- Best, C. A., Yim, H., & Sloutsky, V. M. (2013). The cost of selective attention in category learning: Developmental differences between adults and infants. *Journal of Experimental Child Psychology, 116*, 105–199.
- Deng, S., & Sloutsky, V. M. (2012). Carrot eaters or moving heads: Inductive inference is better supported by salient features than by category labels. *Psychological Science, 23*(2), 178–186.
- Deng, S., & Sloutsky, V. M. (2015). The Development of Categorization: Effects of Classification and Inference Training on Category Representation. *Developmental Psychology, 51*(3), 392–405.
- Hoffman, A. B., & Rehder, B. (2010). The costs of supervised classification: The effect of learning task on conceptual flexibility. *Journal of experimental psychology. General, 139*(2), 319–340. doi: 10.1037/a0019042
- Rabi, R., Miles, S. J., & Minda, J. P. (2015). Learning categories via rules and similarity: Comparing adults and children. *Journal of Experimental Child Psychology, 131*, 149–169.
- Rabi, R., & Minda, J. P. (2014). Rule-based category learning in children: The role of age and executive functioning. *PLoS ONE, 9*(1).
- Rehder, B., & Hoffman, A. B. (2005). Eyetracking and selective attention in category learning. *Cognitive Psychology, 51*(1), 1–41. doi: 10.1016/j.cogpsych.2004.11.001
- Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology, 7*(4), 573–605. doi: 10.1016/0010-0285(75)90024-9
- Schyns, P. G., Goldstone, R. L., & Thibaut, J. P. (1998, feb). The development of features in object concepts. *The Behavioral and brain sciences, 21*(1), 1–54.
- Shepard, R. N., Hovland, C. I., & Jenkins, H. M. (1961). *Learning and memorization of classifications*. (Vol. 75) (No. 13). doi: 10.1037/h0093825
- Sloutsky, V. M. (2010). From Perceptual Categories to Concepts: What Develops? *Cognitive Science, 34*(7), 1244–1286. doi: 10.1111/j.1551-6709.2010.01129.x
- Smith, L. B. (1989). A model of perceptual classification in children and adults. *Psychological review, 96*(1), 125–144. doi: 10.1037/0033-295X.96.1.125
- Smith, L. B., & Kemler, D. G. (1977). Developmental trends in free classification: Evidence for a new conceptualization of perceptual development. *Journal of Experimental Child Psychology, 24*(2), 279–298.