Learning in the wild - how labels influence what we learn

Samuel Rivera (rivera.162@osu.edu)
Department of Psychology, The Ohio State University
1835 Neil Avenue, Columbus, OH 43210 USA

Christopher W. Robinson (robinson.777@osu.edu)
Department of Psychology, The Ohio State University Newark
1179 University Dr., Newark, OH 43055, USA

Abstract

Learning concepts and categories in the real world is often accompanied by verbal labels. The existing theoretical accounts of how labels influence what we learn range from facilitation to overshadowing, with changes occurring over development. Studies investigating how labels influence what people learn have typically been confined to a category learning framework, where participants were tasked to learn how to discriminate categories or infer missing category properties. Here, we investigate how the absence or presence of labels, both common and unique, alter how people attend and what they remember in a more general setting. Our results suggest that unique labels may promote visual exploration of objects; whereas, there was no evidence to support the claim that hearing the same label associated with different members of a to-be-learned category directed attention to common features.

Keywords: categorization; cross-modal processing; attention; learning

Introduction

Every day we learn about objects in the world through experiences with our environment and communication with others. Language, and labeling in particular, is known to have important consequences for how we interpret and perceive categories in our environment (Lupyan, 2012; Perry & Lupyan, 2014), and effects of labeling on visual attention may illuminate a mechanism underlying how language may affect thought more generally (Whorf, 1956). The goal of this research is to understand how linguistic labels, both common and unique, influence what we learn from our environment.

Previous research has shown that adults expertly recognizing the link between labels and concepts (Lupyan, Rakison, & McClelland, 2007; Perry & Lupyan, 2014), and adults appear to appreciate the symbolic nature of linguistic input (Yamauchi & Markman, 2000). Referring to an object with the phrase “it’s a jambo” immediately conveys what the object is, and that it belongs to a particular category of things. The powerful effect of labels exists even in infancy, with studies showing that linguistic labels, unlike non-linguistic sounds, facilitate category formation in very young infants (Althaus & Mareschal, 2014; Balaban & Waxman, 1997; Ferry, Hespos, & Waxman, 2010; but see Robinson & Sloutsky, 2007). The mechanism underlying this ability, especially early in life, remains an active and contentious research topic.

Improved learning has been theorized as stemming from labels indicating category membership in a natural kind class (Gelman & Markman, 1986), and may invite infants to form categories by denoting category membership and highlighting object commonalities (Waxman & Booth, 2003; Waxman & Markow, 1995). It is also possible that hearing different words associated with different objects directs attention to unique features. Support for this claim comes from research examining effects of labels on individuation (Xu, 2002, but see Robinson & Sloutsky, 2008). For example, young infants are more likely to assume that two different objects are hidden behind an occluder if the objects were associated with two unique labels. Infants do not make this assumption if they hear one label associated with both objects or if both objects are associated with two non-linguistic sounds.

The view that labels draw attention to common features and possibly unique features has been supported by behavioral and eye tracking experiments with young infants (Althaus & Plunkett, 2015a; Plunkett, Hu, & Cohen, 2008). However, it is important to note that other studies examining visual attention while learning categories have failed to find evidence that labels direct attention to category relevant features (Best, Robinson, & Sloutsky, 2011). Thus, it is possible that, early in life, labels can affect category learning without being perceived as symbols that denote category membership. They may instead be an additional perceptual cue that adds to the perceptual similarity of objects that are labeled with the same name (Sloutsky & Fisher, 2004; Sloutsky, 2003), but become more like category markers over development (Sloutsky, Lo, & Fisher, 2001).

A third proposal is that labels may actual hinder category learning early in life, with evidence of both words and non-linguistic sounds attenuating both categorization and individuation early in development (Robinson & Sloutsky, 2007, 2008). More specifically, while both of these studies provide some evidence that labels can have facilitative effects compared to non-linguistic sounds, this difference stemmed from non-linguistic sounds interfering with visual processing more than words, as neither the sound nor label condition exceeded a silent condition. This contradictory finding may be reconciled by recent research arguing that the precise synchrony of labels with visual object presentation has important consequences on whether or not labels will facilitate learning (Althaus & Plunket, 2015b; see also Roberts, 1995).

In summary, it has been proposed that labels have a facilitative effect on categorization and individuation in infants and young children by directing attention to common features and unique features, respectively (Waxman & Booth, 2003). However, most of the research supporting the facilitative effects of labels examines the outcome of learning (e.g., better learning in the label condition than a non-linguistics sound condition), and does not directly test the hypothesis that labels affect visual attention. Moreover, the research directly examining infants' and children's fixations while learning categories are mixed (Althaus & Plunkett, 2015b; Best et al., 2011). Therefore, the current study examined the effects of common and unique labels on adults' category learning.

Previous research suggests that for adults, labels are not just perceptual features (Yamauchi & Markman, 2000). Furthermore, adults optimize their attention to focus on category relevant features (Blair, Watson, & Meier, 2009; Hoffman & Rehder, 2010). Will adults who are experts in language learning and attention optimization direct their attention to common features when the to-be-learned individuals are associated with the same labels? Will adults fixate more on unique features when every member of the to-be-learned category is associated with a unique label? To address these questions, we taught adults three novel categories. One category was presented in silence (baseline), members of a second category were always paired with the same word (common label condition) and members of the third category were associated with unique labels (unique label condition). If common labels direct attention to common features, then fixating on common features in the common label condition should exceed the silent baseline. If unique labels direct attention to unique features, then adults in the unique label condition should look less to the common feature and more to unique features compared to the silent baseline.

Method

Participants

Nineteen adults participated in the study for course credit after giving written consent. The study was evaluated and approved by an institutional review board at The Ohio State University.

Stimuli

Verbal labels were pre-recorded sound clips of count nouns in a child directed carrier phrase like "Look, here's a dax" or "this is called a dax" that were spoken by a female. Sound clips were recorded using a Yeti Pro microphone and edited using Audacity software so that each linguistic phrase was approximately one second in duration. Auditory stimuli were presented via Kensington KMW33137 headphones at approximately 65 dB. Visual stimuli were artificial creatures generated with the Spore creature creation software (Electronic Arts Inc., 2009). Creatures were from three different categories defined by the presence of a single deterministic feature. All other features varied independently of the category, and gave no information about category membership. Category A was defined by the presence of three small prongs on the front shoulders. Category B members has a dorsal fin, and category C creatures had suction-cup feet. The new items were perceptually impoverished, had fewer limbs, and were used as catcher items on test trials to make sure participants

Figure 1: We show example stimuli used in the experiment. Each category is defined by the presence of a particular feature: prongs on the shoulders for category A, a dorsal fin for category B, and suction cup feet for category C.

(a) Category A (b) Category B (c) Category C

Figure 2: We show examples of stimuli used in our study, with red circles denoting the areas of interest.

were paying attention during training.

We made an effort to normalize for the overall size and complexity of creatures within and across categories and exemplars. Examples are shown in Figure 1. On average, stimuli spanned approximately 18 by 13 degrees of visual angle and were presented on a 1920 x 1080 Benq XL2420-B monitor. The area of interest (AOI) for each creature was fixed size circle (or circles) enclosing the deterministic feature(s). Some example AOIs are shown in in Figure 2.

Procedure

We conducted a within subject study with eye-tracking to identify differential effects of labeling on attention allocation, memory for encountered objects, and attention to novel exemplars of studied categories. The label conditions encountered during training were Common (single label for category members), Silent (no labels or carrier phrase), or Unique (each category exemplar had a different label). Following training, participants performed an old vs. new recognition task.

To begin, participants were told that they would see a series of creatures from another planet, and that they would learn some of the creatures' names. They were asked to study the creatures at their own pace since they would be asked questions in the second part of the study (testing phase). They were not told about the specific recognition task, nor were they told what they would be tested on (e.g., words, creatures, word-creature pairings), but they were informed about a test in general. Their gaze was recorded during training and testing at 500Hz by an Eyelink 1000 Plus Tower system (SR Research, Ontario, Canada). Fixation information was identified on-line during the experiment by the Eyelink system, then recorded for offline processing and analysis with custom MATLAB and Python software.

Each trial began with a central drift correction target that corrected for slight eye-tracker drift during the experiment. A trained experimenter sitting at the right side of the participant initiated each trial once the participant fixated the central cue. Stimuli were presented in a self-paced manner, in three blocks of 15 trials consisting of 5 items from each category. Participants pressed the 1 key on the number pad when they were ready to move to the next stimulus. Trials within blocks were randomized, and the same 15 items were presented in each block. Each category was consistently presented in either silence (category A), with a unique label phrase (category B), or with a common label phrase (category C). The auditory and visual stimuli shared the same stimulus onset. The carrier phrase and label terminated after about 1 s and the visual stimulus terminated when the participant pressed the 1 key.

After three training blocks, participants were presented with an old versus new recognition task. All of these test trials were presented in silence. Participants saw 5 exemplars from each of the three categories that had been previously studied (15 old items), 5 novel exemplars from the studied categories (15 lure items that had the category defining feature), and 6 completely new trials that looked completely different from any of the studied items. Each trial began with a drift correction target as before. Participants responded old or new to the stimulus by pressing the 1 key for old, and the 3 key for new. Participants were instructed to respond as fast and as accurately as possible.

If common labels facilitate categorization by directing attention to common features, then participants in the common label condition should focus more on the category defining features compared to the silent condition, and thus would be more likely to mistake critical lures as old or have worse recall in general in comparison to the unique label or silent conditions. If unique labels facilitate individuation by directing attention away from common features and to unique features, then participants in the unique label condition should focus more on unique features compared to the silent condition.

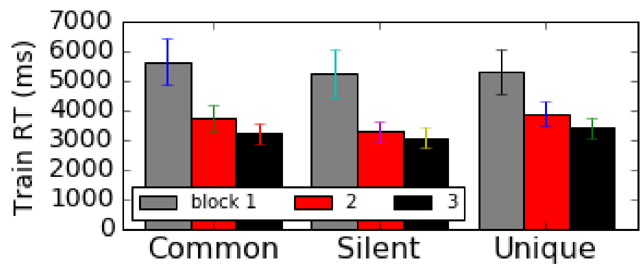

Figure 3: Here we show participant mean study time across conditions and blocks. Error bars denote standard error. Trends show that participants spent less time studying over the training blocks.

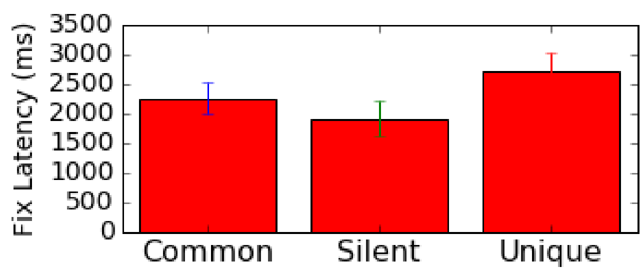

Figure 4: Here we show participant mean latency to fixate the deterministic feature. Error bars denote standard error.

Results

Training Results

Training study time We plot the mean study time in Figure 3. We ran a 3 (condition: common vs. silent vs. unique) x 3 (block) repeated measures ANOVA on log(studyTime). We found a significant effect of condition, $F(2,36) = 9.56, p < .001, \eta_p^2 = 0.35$. Pairwise comparisons showed that participants viewed the images for less time in the Silent condition than the Common ($p = 0.015$, Bonferroni corrected) and the Unique label conditions ($p < .001$). Thus, participants were faster when only the visual information was presented. There was also a main effect of block, $\text{F}(2,36) = 31.05, p < 0.001, \eta_p^2 = 0.63$. Pairwise comparison also showed that block 1 was significantly longer than blocks 2 and 3 ($ps < 0.001$). Thus, participants spent less time studying the creatures after the first block.

Training latency Deterministic Looks For several participants, there were blocks where the deterministic features were never fixated. Thus, we excluded these participants from this analysis. We considered the averages across all blocks such that only one participant was excluded. Averages are shown in Figure 4. We ran a one-way repeated measures ANOVA on latency to fixate the relevant AOI with main effect of condition (common vs. silent vs. unique). We found a significant main effect of condition, $F(2,34) = 10.78, p < .001, \eta_p^2 = .39$. Pairwise comparisons revealed that only the

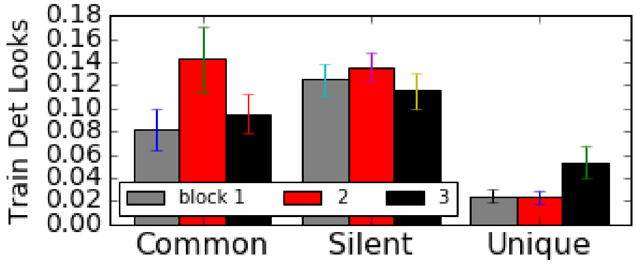

Figure 5: Here we show participant mean proportion of looking at the deterministic or category defining AOI. Error bars denote standard error.

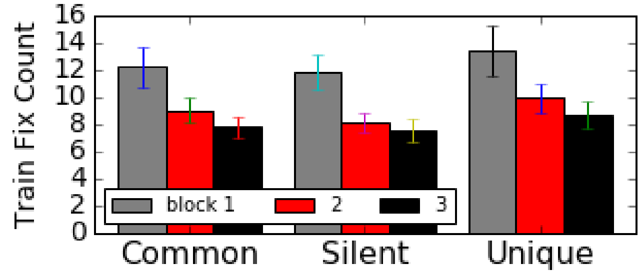

Figure 6: Here we show participant mean number of fixations during training across conditions and blocks. Error bars denote standard error.

difference between the silent and unique conditions was statistically significant ($p < 0.001$, Bonferroni corrected), with latency in the silent condition being shorter. The difference between the silent and common was approaching significance ($p = 0.077$), with latency in the silent condition being shorter than the common label condition. Thus, participants in the silent condition fixated the deterministic feature earlier on trials during training. Is it possible that participants were in general fixating earlier on Silent trials? We ran a 3 (condition: common vs. silent vs. unique) x 3 (block) repeated measures ANOVA on latency of log(first fixation) to test this possibility. There were no significant main effects or interactions, ruling out the possibility that participants in the silent condition may have been faster overall.

Training - deterministic looking The latency analysis above is problematic because there were many trials where the deterministic feature was never fixated so that latency information was missing. Such cases bias the result because if the deterministic feature was never fixated, that trial was excluded from the analysis instead of providing a large latency value. A better measure, if we are interested in how label conditions contribute to deterministic looking is to consider the proportion of looking on a trial. Thus, no fixations at the deterministic feature would give a value of 0. We show the averages across conditions and blocks in Figure 5. We ran a 3 (condition: common vs. silent vs. unique) x 3 (block) repeated measures ANOVA on proportion of fixation time at the deterministic feature. We found a significant main effect of condition, $F(1.65, 29.78) = 21.39, p < 0.001, \eta_p^2 = 0.54$, Greenhouse-Geisser corrections. Pairwise comparisons revealed that the unique condition had a significantly lower proportion of fixation time at the deterministic feature than the other conditions ($ps < 0.003$, Bonferroni corrected). No other main effects or interactions were significant.

Training Fixation counts Given this difference in proportion of fixations at the deterministic feature, we wondered if this could be explained by an overall increase in the exploration of the object. Such behavior would transpire in an increase in the number of overall fixations. We plot means in Figure 6. A 3 (condition: common vs. silent vs. unique) x 3 (block) repeated measures ANOVA on number of fixations revealed a statistically significant main effect of condition, $F(2, 36) = 5.66, p = 0.007, \eta_p^2 = 0.24$. Pairwise comparisons showed that the effect was driven by a significant difference between the silent and unique condition (p=0.02, Bonferroni corrected), with more fixations in the unique condition as predicted. The number of fixations in the unique condition was trending larger than the common label condition, but the difference was not statistically significant.

There was also a significant main effect of block, $F(1.31, 23.54) = 13.63, p = 0.001, \eta_p^2 = 0.43$, Greenhouse-Geisser corrected for sphericity. Pairwise comparisons revealed that there were significantly more fixations in the first block than later blocks ($ps < 0.03$, Bonferroni corrected). Interaction between condition and block was not significant.

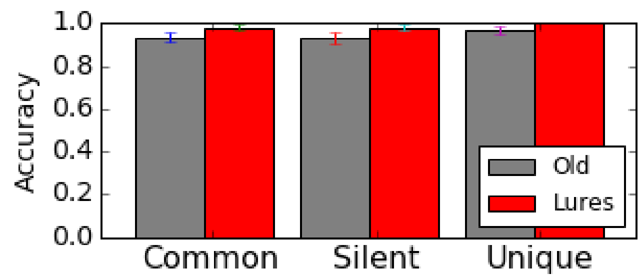

Figure 7: Here we show participant mean accuracy across conditions for both old and lure trials. Error bars denote standard error. Trends suggest that lure items were more accurately identified as new across all conditions. However the differences were not significant.

Test Results

Accuracy One participant reversed the response mapping, as indicated by 0% accuracy. We reversed their responses in calculating the accuracy scores below. We plot the mean participant accuracy in Figure 7. Overall, participant accuracy was near or at ceiling in the old vs. new judgement task, with accuracy above 95% across the three conditions. A 3 (condition: common vs. silent vs. unique) x 2 (test type: Lure

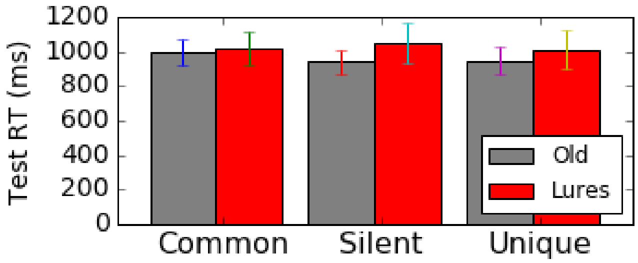

Figure 8: Here we show mean participant reaction times during test trials. Error bars denote standard error.

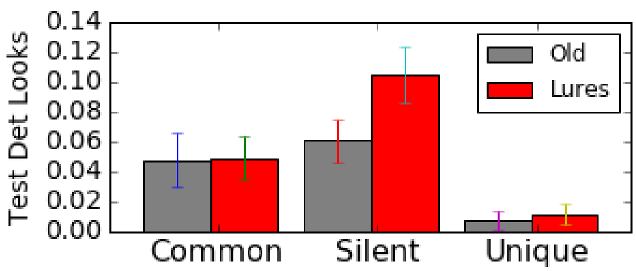

Figure 9: Here we show mean proportion of fixation time at the deterministic or category defining feature. Error bars denote standard error.

vs. Old) repeated measures ANOVA on old/new recall accuracy revealed a statistically significant main effect of test type, $F(1,17) = 5.64, p = 0.03, \eta_p^2 = 0.25$, with accuracy on lure items being higher. No other effects or interactions were significant.

Testing RT Mean reaction times during test trials are shown in Figure 8. We applied a 3 (condition: common vs. silent vs. unique) x 2 (test type: Lure vs. Old) repeated measures ANOVA to log(RT) on accurate test trials. There were no significant main effects or interactions.

Testing proportion of deterministic looking We plot the proportion of fixation time at the deterministic AOI in Figure 9. A 3 (condition: common vs. silent vs. unique) x 2 (test type: Lure vs. Old) repeated measures ANOVA on proportion of deterministic looking on test trials revealed a significant main effect of condition, $F(2,34) = 13.16, p < 0.001, \eta_p^2 = 0.44$. Pairwise comparisons showed that the effect was driven by the unique condition having a significantly lower proportion than the common label condition ($p = 0.031$, Bonferroni corrected) and the silent condition ($p < 0.001$). This result is consistent with the gaze pattern during training.

Testing Fixation counts As with the training data, we were interested in whether the relatively smaller proportion of fixations at the deterministic feature could be explained by an overall increase in the number of fixations in the unique condition. We show the average number of fixations in Figure 10.

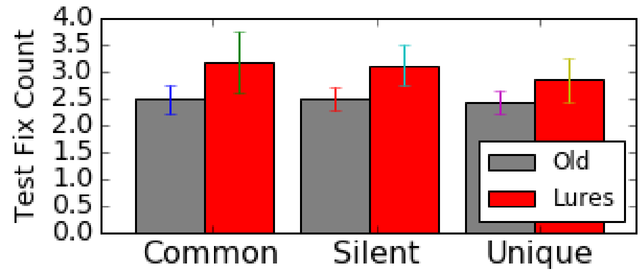

Figure 10: Here we show mean number of fixations during testing trials. Error bars denote standard error.

A 3 (condition: common vs. silent vs. unique) x 2 (test type: Lure vs. Old) repeated measures ANOVA on number of fixations revealed no significant main effects or interactions. Thus, across conditions participants fixated the same amount when making their old/new judgements. However, the unique label condition was associated with less focus on the deterministic feature(s).

Discussion

In the current study we investigated how the presence or absence of labels affect adults' visual attention while viewing novel visual stimuli. Previous research with infants and young children shows that infants are more likely to learn categories when the to-be-categorized images are paired with the same label (Balaban & Waxman, 1997) and more likely to individuate exemplars or learn two categories when the different objects are associated with different labels (Plunkett et al., 2008; Xu, 2002). One potential mechanism that may account for these effects is that infants understand the symbolic nature of language, with common and unique words (but not sounds) directing infants attention to common and unique features, respectively (Waxman & Booth, 2003). However, recent eye tracking studies directly testing this hypothesis have yielded mixed results (Althaus & Plunkett, 2015a; Best et al., 2011); therefore, this study examined visual fixations in adults, who clearly understand the symbolic nature of words and can optimize their attention (Blair et al., 2009; Lupyan et al., 2007; Rehder & Hoffman, 2005).

Participants in the current study required less exposure to images and were faster to make fixations to category defining features when the images were presented in silence. This is consistent with some of the research examining auditory dominance in a variety of cognitive tasks (see Robinson et al. 2012 for a review). Across many different behavioral and eye tracking variables, there was no support for the hypothesis that common labels directed attention to category relevant features; however, there was some support showing that hearing unique labels associated with category members decreased attention to commonalities or increased attention to unique features.

A goal of future research is to adapt this initial study to examine labeling effects in younger populations. One advantage of the study is that it mirrors situations encounter outside the laboratory, where the way knowledge will be applied is unknown. However, this article is limited in that stimuli are defined by only a small subset of defining features, and the defining features did not vary across labeling phrases. Thus, future work should consider categories comprised of multiple defining features with proper counter-balancing of these categories with the labeling conditions.

In summary, studies have demonstrated important influences of labeling on category learning. While many of these studies show that labels and sounds can have different effects on categorization and individuation, underlying mechanism are poorly understood. The current study took a novel approach by examining fixations in real time as adults learned about category objects. While there was some support showing that unique labels push attention to unique features or away from category defining features, there was no support showing that participants looked more at defining features or made faster fixations to these defining features. These findings have implications for proposed mechanisms underlying the effects of labels on category learning and should be further examined in younger populations.

References

Althaus, N., & Mareschal, D. (2014). Labels direct infants' attention to commonalities during novel category learning. *PLoS ONE*, *9*(7).

Althaus, N., & Plunkett, K. (2015a). Categorization in infancy: Labeling induces a persisting focus on commonalities. *Developmental Science*.

Althaus, N., & Plunkett, K. (2015b). Timing matters: The impact of label synchrony on infant categorisation. *Cognition*, *139*, 1–9.

Balaban, M. T., & Waxman, S. R. (1997). Do words facilitate object categorization in 9-month old infants? *Journal of Experimental Child Psychology*, *64*, 3–26.

Best, C. A., Robinson, C. W., & Sloutsky, V. M. (2011). The effect of labels on categorization: Is attention to relevant features a good index of infants category learning? In L. Carlson, C. Hlscher, & T. Shipley (Eds.), *Proceedings of the 33rd annual conference of the cognitive science society* (pp. 2751–2755).

Blair, M. R., Watson, M. R., & Meier, K. M. (2009). Errors, efficiency, and the interplay between attention and category learning.

Ferry, A., Hespos, S., & Waxman, S. (2010). Categorization in 3- and 4-month-old infants: An advantage of words over tones. *Child Development*(2), 472–479.

Gelman, S. a., & Markman, E. M. (1986). Categories and induction in young children. *Cognition*, *23*(3), 183–209.

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.

Lupyan, G. (2012). Linguistically modulated perception and cognition: the label feedback hypothesis. *Frontiers in Psychology*, *3*, 54.

Lupyan, G., Rakison, D. H., & McClelland, J. L. (2007). Language is not just for talking: Redundant labels facilitate learning of novel categories. *Psychological Science*, *18*(12), 1077–1083.

Perry, L. K., & Lupyan, G. (2014). The role of language in multi-dimensional categorization: Evidence from transcranial direct current stimulation and exposure to verbal labels. *Brain and Language*, *135*, 66 - 72.

Plunkett, K., Hu, J. F., & Cohen, L. B. (2008). Labels can override perceptual categories in early infancy. *Cognition*, *106*, 665–681.

Rehder, B., & Hoffman, A. B. (2005). Eyetracking and selective attention in category learning. *Cognitive Psychology*, *51*(1), 1–41.

Roberts, K. (1995). Categorical responding in 15-month-olds: Influence of the noun-category bias and the covariation between visual fixation and auditory input. *Cognitive Development*, *10*, 21–41.

Robinson, C. W., Best, C. A., Deng, S., & Sloutsky, V. M. (2012). The role of words in cognitive tasks: What, when, and how? *Frontiers in Psychology*, *3*, 1–8.

Robinson, C. W., & Sloutsky, V. M. (2007). Linguistic labels and categorization in infancy: Do labels facilitate or hinder? *Infancy*, *11*(3), 233–253.

Robinson, C. W., & Sloutsky, V. M. (2008). Effects of auditory input in individuation tasks. *Developmental Science*, *11*, 869–881.

Sloutsky, V. M. (2003). The role of similarity in the development of categorization. *Trends in Cognitive Sciences*, *7*(6), 246 –251.

Sloutsky, V. M., & Fisher, A. V. (2004). Induction and categorization in young children: a similarity-based model. *Journal of experimental psychology. General*, *133*(2), 166–188.

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.

Waxman, S. R., & Booth, A. E. (2003). The origins and evolution of links between word learning and conceptual organization: New evidence from 11-month-olds. *Developmental Science*, *6*, 130–137.

Waxman, S. R., & Markow, D. B. (1995). Words as invitations to form categories: evidence from 12- to 13-month-old infants. *Cognitive psychology*, *29*(3), 257–302.

Whorf, B. (1956). *Language, thought, and reality* (J. B. Carroll, Ed.). Cambridge, Massachusetts: MIT Press.

Xu, F. (2002). The role of language in acquiring object kind concepts in infancy. *Cognition*, *85*, 223–250.

Yamauchi, T., & Markman, A. B. (2000). Inference using categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 776–795.