Title
The Effect of Labels on Visual Attention: An Eye Tracking Study

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 32(32)

ISSN
1069-7977

Authors
Best, Catherine
Robinson, Christopher
Sloutsky, Vladimir

Publication Date
2010

Peer reviewed

The Effect of Labels on Visual Attention: An Eye Tracking Study

Catherine A. Best (best.140@osu.edu)
Center for Cognitive Science, The Ohio State University
208G Ohio Stadium East, 1961 Tuttle Park Place, Columbus, OH 43210, USA

Christopher W. Robinson (robinson.777@osu.edu)
Center for Cognitive Science, The Ohio State University
208F Ohio Stadium East, 1961 Tuttle Park Place, Columbus, OH 43210, USA

Vladimir M. Sloutsky (sloutsky.1@osu.edu)
Center for Cognitive Science, The Ohio State University
208D Ohio Stadium East, 1961 Tuttle Park Place, Columbus, OH 43210, USA

Abstract

The effects of language on categorization are well documented; however, underlying mechanisms are under debate. According to one account, words facilitate categorization by highlighting commonalities among labeled objects. Although there is some behavioral evidence consistent with this claim, research remains limited for whether labels can direct infants' attention to corresponding visual features. In the current study, adults and infants were presented with 10 different exemplars that were either associated with 10 different labels, the same label, or presented in silence. An eye tracker recorded visual fixations to common and unique features throughout familiarization. Experiments 1 and 2 provide evidence that unique labels can direct infants' and adults' attention to unique features (compared to a silent condition); however, the effect of hearing the same label associated with different objects was less robust in both age groups.

Keywords: Attention; Language; Categorization

Introduction

Beginning at birth, infants must learn to make sense of the world, and the ability to form categories is an important part of this learning. Although very young infants can quickly learn visual categories (Bomba & Siqueland, 1983; Eimas & Quinn, 1994), there is some evidence that words and other types of sounds influence this process. For example, young infants are often better at learning visual categories when category members are associated with the same word than when the same visual stimuli are paired with a nonlinguistic sound (Balaban & Waxman, 1997; Fulkerson & Waxman, 2007; Robinson & Sloutsky, 2007). Exposure to words may also help infants individuate objects. Research demonstrates that infants who hear two different words (but not two sounds) expect two objects to be hidden by an occluder (Xu, 2002). Labels also influence what category structure infants learn. For instance, while looking at the same visual images, infants who heard one word associated with all exemplars learned one category; whereas, infants who heard two words learned two categories (Plunkett, Hu & Cohen, 2008). Finally, although words and sounds often have different effects on categorization and individuation, only a few studies have directly compared infants' performance in label and sound conditions to a silent baseline. These comparisons illustrate that compared to a silent condition, words and sounds can interfere with categorization of visual input, often with greater interference from sounds than from words (Robinson & Sloutsky, 2007; 2008).

To account for the effect of labels on category learning, several mechanisms have been put forth. First, Waxman and colleagues argue that infants understand the conceptual importance of words and that words (but not sounds) facilitate categorization by highlighting the commonalities among labeled entities (Fulkerson & Waxman, 2007; Waxman, 2003). Given the findings reported by Plunkett *et al*. (2008) and Xu (2002), it is also possible that unique words may also facilitate the formation of multiple categories by highlighting unique features among labeled entities. In contrast, Sloutsky and colleagues argue that infants and young children have difficulty processing multimodal information, with words and sounds often attenuating visual processing (Robinson & Sloutsky, 2004; Sloutsky & Napolitano, 2003). Differential effects of words and sounds stem from sounds interfering with visual tasks more than words (as opposed to words facilitating categorization above a silent control). Thus, according to Waxman and colleagues, hearing common and unique words should increase attention to common and unique features in the early stages of development. In contrast, according to Sloutsky and colleagues, early in development words should have no facilitative effect above a silent condition and may even interfere with visual processing.

The aim of the current set of studies was to explore *how* words might affect visual attention by utilizing eye-tracking technology. Measuring eye movements during experimental tasks provides an online measure of attention. By tracking the gaze of infants and adults during a simple familiarization task, we can investigate whether patterns of visual attention during learning differ with respect to varying language cues.

Overview of Current Studies

To investigate the effect of labels on visual attention, gaze data were collected from both infants and adults while viewing novel stimuli paired with novel labels. Half the features on each stimulus were shared among the sequentially presented stimuli (i.e., common features); whereas, half of the features were unique. If participants inferred identical labels indicated that images were members of the same object category, it was predicted that participants who heard the same label associated with different images would accumulate more looking to common features than participants in the silent condition. Similarly, if participants inferred different labels indicated that images were members of different object categories, it was predicted that participants who heard different labels associated with different images would accumulate more looking to unique features than participants in the silent condition. Experiment 1 compared adults' attention to common and unique features across familiarization when labels were consistent, varying, or when images were presented in silence. Experiment 2 tested infants with the same three sets of stimuli as presented to adults.

Experiment 1

Method

Participants Thirty-six adults (20 men, 16 women), ranging in age from 18 to 21 years (M = 18.58, SD = 0.79) were tested, with 12 adults per condition. Adults were recruited from an Introductory Psychology class. Participants provided written consent upon arrival to the laboratory. All adults reported normal or corrected-to-normal vision and normal hearing prior to recruitment.

Apparatus A non-invasive Tobii T60 eye tracker measured eye gaze by computing the pupil-corneal reflection at a sampling rate of 60 Hz (i.e., 60 gaze data points collected per second for each eye). The eye-tracking device, which is integrated into the base of a high-resolution 17-inch computer monitor, was located on a table inside a darkened testing booth, enclosed by curtains. A trained experimenter monitored the experiment on a 19-inch Dell OptiPlex 755 computer located outside of the testing booth. A Sony Network camera was located inside the testing booth to the side of the eye tracker displaying a live feed view of the participant that an experimenter monitored on a 9-inch black and white Sony SSM-930/930 CE television. Two Dell computer speakers were positioned behind a curtain and out of view on either side of the eye tracker.

Stimuli Stimuli included 12 audio-video interleave (AVI) files. Each AVI file combined a static bitmap image with an auditory speech component. The visual images consisted of four uniquely-shaped parts. Two parts were common across all stimuli and two parts were unique across all stimuli. See Figure 1 for example stimuli. The common parts were the same color and shape; whereas, the unique parts varied in color and shape. The auditory input consisted of one-syllable novel labels spoken by a female adult (e.g., *dax, bim*, *fep*, *gid*, *jup*, *meb*, *pof*, *raz*, *sop*, and *zot*). All labels were spoken within the context of a simple command (*e.g.,* "Look at the dax."). Speech was recorded using Cool Edit 2000. Each sound file was saved as an audio compression manager waveform at 44.10 kHz, 16 Bit, in stereo. Audio files were then imported into Macromedia Flash, paired with corresponding bitmap images, and exported as Windows AVI files. All AVI files were 6000 ms in duration. The image lasted for the total duration of 6000 ms. The audio occurred with the onset of the image and lasted 2000 ms in duration. The remaining 4000 ms consisted of silence. The stimuli for the silent condition were identical to the label conditions; except, the speech was removed entirely.

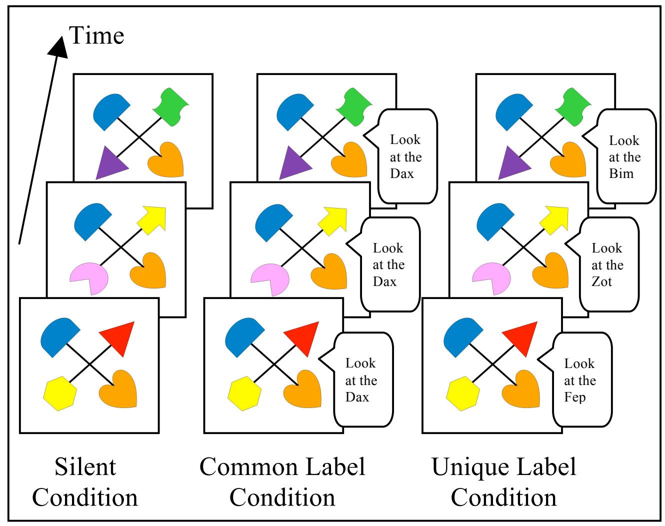

Figure 1: Example stimuli

Design The experiment utilized a between-subjects design. Participants were randomly assigned to one of three experimental conditions (*i.e*., common label, unique label, or silent). The common condition consisted of 10 different novel images, each paired with the same novel label. The unique condition consisted of 10 different novel images, each paired with 10 different novel labels. The silent condition consisted of 10 different novel images, each presented in silence. The visual input was the same for all conditions and was presented in a random sequence. Only the auditory input differed across conditions.

Procedure Participants sat centered in front of the eye tracker within an approximate viewing distance of 60 cm. Prior to the experiment, participants completed a 5-point calibration sequence lasting less than one minute. The calibration points consisted of a moving red dot appearing in different locations on the screen. The experiment commenced after successful calibration. Participants were asked to pay close attention to the images because they would be asked about them later. All participants were familiarized to 10 stimuli presented one at a time for 6000 ms. Each stimulus was presented subtending an approximate horizontal visual angle of 11° and an approximate vertical visual angle of 11°. A black screen was presented for 1000ms between trials. After training, participants were tested with four paired preference trials, each trial displaying one old image and one new image presented in silence. Adults were asked to select the new image from the old image. Test stimuli were the same size as familiarization stimuli. Each test trial remained visible until adults made a decision. All gaze data were recorded by the computer using Tobii Studio gaze analysis software.

Results and Discussion

All participants correctly identified the novel stimuli on every test trial; therefore, no one was excluded from the current study. Primary analyses presented below focused on adults' attention to common and unique features during familiarization.

Unfiltered gaze data were exported from the computer using Tobii Studio gaze analysis software. A point of gaze was recorded if a participant made a fixation to pre-determined areas of interests (AOIs). Four AOIs were defined as a rectangle surrounding the four parts of each stimulus image. Gaze data were combined for the two common features and for the two unique features to obtain a measure of looking to unique or common features per refresh rate. These data were used to calculate unique and common feature preference scores based on the proportion of looking time to either unique or common features as compared to the total time looking to all features combined.

Effect of Unique Labels To determine if unique labels pushed adults' attention to unique features, we compared preference for unique features in the unique label condition to preference for unique features in the silent condition.

Gaze data were analyzed using a moving average of participants' attention across time to smooth out temporary fluctuations within a given trial (*i.e.*, 3 trials were averaged per time point such that time point 1 averaged trials 1 to 3, time point 2 averaged trials 2 to 4, and so on). A repeated-measures analysis of variance (ANOVA) was conducted on mean preference for unique features with Condition (unique label vs. silent) as a between-subjects factor and Time Point (1 vs. 2 vs. 3 vs. 4 vs. 5 vs. 6 vs. 7 vs. 8) as a within-subjects factor. Results revealed a significant main effect of Time Point, $F(7, 154) = 6.56$, $p < .001$, a main effect of Condition, $F(1, 22) = 4.44$, $p < .05$, and a significant Time Point X Condition interaction, $F(7, 154) = 3.74$, $p < .01$. Overall, mean preference scores for unique features were significantly greater in the unique condition ($M = .66$) compared to the silent condition ($M = .53$). Specifically, as shown in Figure 2, post-hoc comparisons revealed that mean preference scores for unique features were significantly greater in the unique condition compared to the silent condition at time points 5, 6, and 7, ts > 2.16, ps < .05. These results support the idea that unique labels facilitate attention to unique features.

To obtain a better understanding of the dynamics of attention, unique preference scores were averaged across trials and plotted as a function of time. As can be seen in Figure 3, preference for the unique features in the unique label condition was consistent across the entire 6000 ms trial duration. This attention pattern in adults corroborates evidence for unique labels facilitating attention to unique features.

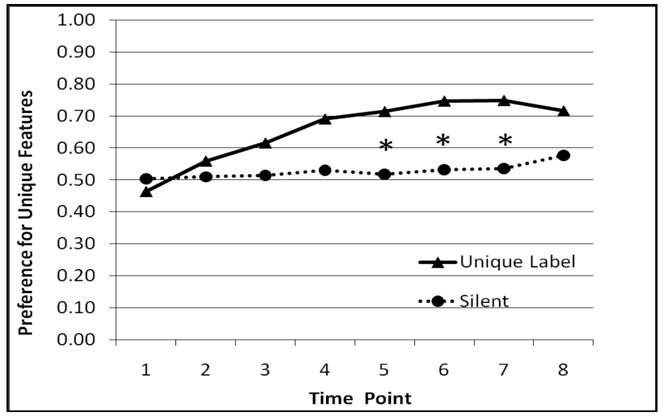

Figure 2: Adults' mean preference for unique features by time point. (Note: Time points represent moving averages).

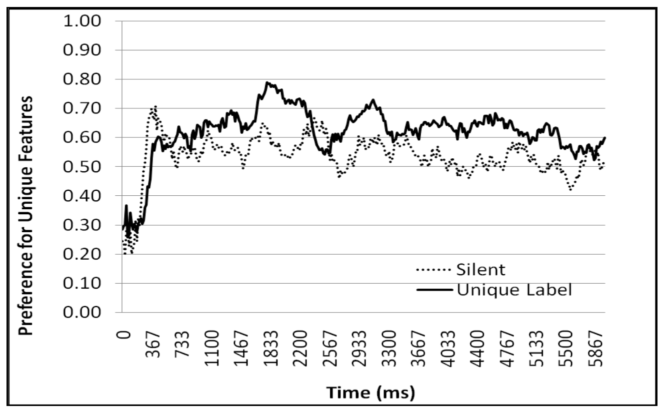

Figure 3: Adults' mean preference for unique features over time.

Effect of Common Labels To determine if common labels pushed adults' attention to common features, we compared preference for common features in the common label condition to preference for common features in the silent condition, using the same sample of adults that was previously compared to the unique label condition.

As in the unique label condition, gaze data were analyzed using a moving average. A repeated-measures ANOVA was conducted on mean preference for common features with Condition (common label vs. silent) as a between-subjects factor and Time Point (1 vs. 2 vs. 3 vs. 4 vs. 5 vs. 6 vs. 7 vs. 8) as a within-subjects factor. Results revealed a significant main effect of Time Point, $F(7, 154) = 2.20$, $p < .05$. Preference for common features attenuated over time for both the common label and silent conditions. However, as shown in Figure 4, mean preference scores for common features were not significantly different between conditions at any point in the course of familiarization.

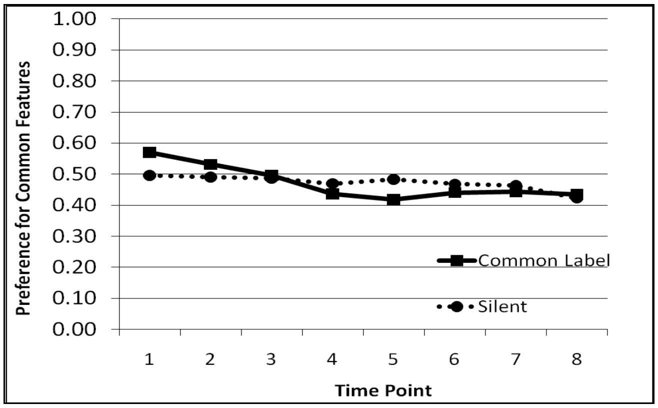

Figure 4: Adults' mean preference for common features by time point. (Note: Time points represent moving averages).

Furthermore, preference for common features was analogous for the entire 6000 ms trial duration in the common label and silent conditions when preferences scores were averaged across trials and plotted as a function of time (see Figure 5). Therefore, the pattern of attention over time did not suggest that common labels directed adults' attention to common features.

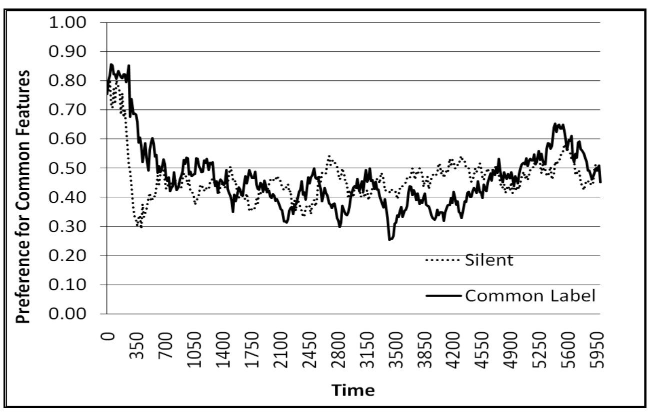

Figure 5: Adults' mean preference for common features over time.

Summary Experiment 1 found a robust effect of unique labels directing attention to unique features and no significant effect of common labels directing attention to common features. Adults presented with varying labels (*i.e.*, unique) compared to silence disproportionately distributed their attention to objects' unique versus common features. In contrast, adults presented with consistent labels (*i.e.*, common) compared to silence did not disproportionately distribute attention to objects' common versus unique features. The purpose of Experiment 2 was to investigate how labels affect visual attention in infancy. Do unique and common labels direct infants' attention to correlated visual features?

Experiment 2

Method

Participants Thirty-six infants, (19 boys and 17 girls), ranging in age from 16 to 24 months (M = 19 months, 9 days; SD = 3 months, 21 days) were tested, with 12 infants per condition. Ten additional infants were excluded from analyses due to fussiness. Infants were recruited from local birth records. Parents provided written consent upon arrival to the laboratory. All infants were healthy and developing typically.

Materials and Design The apparatus, stimuli, and design were identical to Experiment 1.

Procedure Infants sat on a caregiver's lap and were positioned in front of the eye tracker within an approximate viewing distance of 60 cm. The procedure was identical to Experiment 1 with three exceptions. First, during calibration, rather than a shrinking red dot, infants saw a dynamic kitten image appearing on the screen with a corresponding "bounce" sound. Second, unlike adults, infants were not provided with instructions. Third, a dynamic bouncing ball was presented as an attention-grabbing fixation between trials.

Results and Discussion

Infants in all three conditions (*i.e.*, unique label, common label, and silent) demonstrated a mean novelty preference based on the average looking time to new versus old objects across all four test trials, ts > 2.55, ps < .05. Mean novelty preference scores did not differ among the three conditions. Primary analyses presented below focused on infants' attention to common and unique features during familiarization.

Effect of Unique Labels As in Experiment 1, unfiltered gaze data were exported and combined into looking to common features and looking to unique features. To determine if unique labels directed infants' attention to unique features, we compared preference for unique features in the unique label condition to preference for unique features in the silent condition. As with adults' data, infants' gaze data were analyzed using a moving average of participants' attention across time to smooth out temporary fluctuations within a given trial (*i.e.*, 3 trials were averaged per time point such that time point 1 averaged trials 1 to 3, time point 2 averaged trials 2 to 4, and so on). A repeated-measures ANOVA was conducted on mean preference for unique features with Condition (unique label vs. silent) as a between-subjects factor and Time Point (1 vs. 2 vs. 3 vs. 4 vs. 5 vs. 6 vs. 7 vs. 8) as a within-subjects factor. Results revealed a significant main effect of Time Point, $F(7, 147) = 3.96$, $p < .01$. Although the effect of Condition did not reach significance, as shown in Figure 6, independent t-tests revealed that mean unique preference scores were significantly greater in the unique label condition compared to the silent condition at time point 2, $t(22) = 2.03$, $p = .05$.

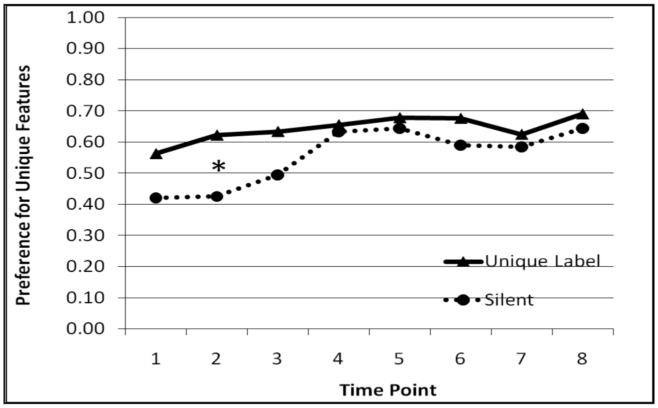

Figure 6: Infants' mean preference for unique features by time point. (Note: Time points represent moving averages).

To obtain a better understanding of the dynamics of attention, unique preference scores were averaged across trials and plotted as a function of time. As can be seen in Figure 7, preference for the unique features was greater in the unique label condition than the silent condition within 1000 ms to 4000 ms. Although, the effect of unique labels was less pronounced in infants than adults, these data provide some evidence for unique labels facilitating infants' attention to unique features.

Figure 7: Infants' mean preference for unique features over time.

Effect of Common Labels To determine if common labels directed infants' attention to common features, we compared preference for common features in the common label condition to preference for common features in the silent condition, using the same sample of infants that was previously compared to the unique label condition. As in the unique label condition, gaze data were analyzed using a moving average. A repeated-measures ANOVA was conducted on mean preference for common features with Condition (common label vs. silent) as a between-subjects factor and Time Point (1 vs. 2 vs. 3 vs. 4 vs. 5 vs. 6 vs. 7 vs. 8) as a within-subjects factor. Results revealed a significant main effect of Time Point, $F(7, 147) = 9.06$, $p < .001$. Like adults, infants' preference for common features attenuated over time for both the common label and silent conditions. However, as shown in Figure 8, mean preference scores for common features were not significantly different between conditions at any point in the course of familiarization.

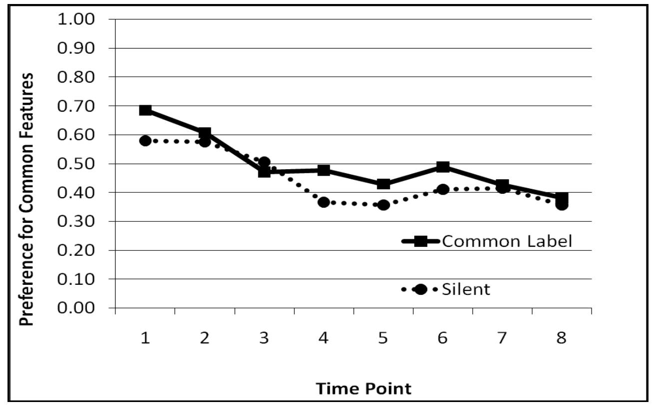

Figure 8: Infants' mean preference for common features by time point. (Note: Time points represent moving averages).

Common preference scores were averaged across trials and plotted as a function of time (see Figure 9). Although results from the ANOVA and t-tests revealed no differences between conditions, preference for the common features in the common label condition exceeded the silent condition for the first and last 1000 ms of the trials. Although not illustrated by adults, this pattern of results revealed that if common labels directed infants' attention to common features, the effects were subtle.

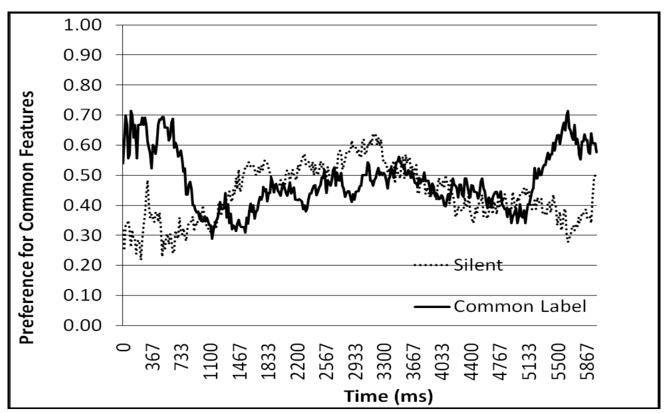

Figure 9: Infants' mean preference for unique features over time.

Summary Experiment 2 found comparable, yet less pronounced results as Experiment 1 with regard to unique labels affecting visual attention. Infants presented with varying labels (*i.e.*, unique) compared to silence disproportionately distributed their attention to objects' unique versus common features. Effects of common words were less robust, and if they directed infants' attention to common features, these effects are subtle.

Conclusion

The current study reveals several important findings. First, adults, and to a lesser extent, infants, who heard unique labels accumulated more looking to unique features compared to the silent condition. Second, for adults, this effect was robust across familiarization and occurred throughout the entire trial. Third, there was no clear evidence of common labels directing attention to common features for adults or for infants.

Many studies have examined how different types of auditory input affect categorization, as assessed by increased looking to novel categories in a subsequent testing phase (Balaban & Waxman, 1997; Fulkerson & Waxman, 2007; Plunkett, Hu & Cohen, 2008; Robinson & Sloutsky, 2007). The current study, in conjunction with research by Althaus and Mareschal (2010), are the first studies we are aware of that have directly tested the hypothesis that words draw attention to category relevant features for infants. Directly testing this hypothesis (*i.e.*, as opposed to inferring it from infants' looking to the novel category at test) is crucial for understanding possible mechanisms underlying effects of words on category learning.

The findings of the current study are partially consistent with both proposed mechanisms. First, in support of the claim that words direct attention to category relevant information (e.g., Waxman, 2003), there was clear evidence for adults, and to a lesser extent, infants, that unique labels highlight unique features. However, there was little support for the claim that common words highlight commonalties, which may have stemmed from a general tendency to increase looking to novel features across familiarization.

Support for the claim that auditory information can attenuate visual processing (e.g., Robinson & Sloutsky, 2007) is also mixed. Support for this claim primarily comes from the finding that infants in the label conditions did not show better discrimination at test, and there was no robust facilitation across familiarization. However, this account assumes that differential effects of words and sounds stem from sounds attenuating visual processing more than words. A direct test of this account would require a non-linguistic sound condition. At the same time, there was little evidence that words slowed down visual processing. However, studies showing that words interfered with visual processing tested 8- and 12-month-old infants (Robinson & Sloutsky, 2007; Sloutsky & Robinson, 2008), which is younger than the infants tested in the current study.

The current study raises an interesting question. Why was the effect of common labels weaker than the effect of unique labels? One possibility is that adults were told to pay attention to the pictures because they were going to be asked about them later. These instructions, in combination with habituation to the common features may have biased attention to unique features. Future research will need to systematically manipulate the category structure by changing the proportion of common to unique features. It is possible that effects of words may interact with the structure of the to-be-learned category. It will also be important to test categorization abilities to connect performance at test to training data, allowing for a better examination of individual differences in category learning.

Acknowledgments

This research is supported by grants from NSF (BCS-0720135), NIH (R01HD056105), and the US Department of Education (R305H050125 and R305B070407) to V. M. Sloutsky and from NIH (RO3HD055527) to C. W. Robinson.

References

Althaus, N., & Mareschal, D. (2010). Speech facilitates categorization but only novel labels direct infants' attention to commonalities. Poster presented at the biennial meeting of International Conference on Infant Studies, Baltimore, MD.

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.

Bomba P.C., & Siqueland E.R. (1983) The nature and structure of infant form categories. *Journal of Experimental Child Psychology, 35*, 294– 328.

Eimas, P., & Quinn, P. (1994). Studies on the formation of perceptually based basic-level categories in young infants. Child Development, 65, 903-917.

Fulkerson, A.L., & Waxman, S.R. (2007). Words (but not tones) facilitate object categorization: evidence from 6- and 12-month-olds. *Cognition, 105*, 218–228.

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

Robinson, C.W., & Sloutsky, V.M. (2004). Auditory Dominance and its change in the course of development. *Child Development, 75 (5), 1387-1401*.

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

Sloutsky, V.M., & Napolitano, A.C. (2003). Is a picture worth a thousand words? Preference for auditory modality in young children. *Child Development, 74*, 822–833.

Sloutsky, V.M., & Robinson, C.W. (2008). The role of words and sounds in visual processing: from overshadowing to attentional tuning. *Cognitive Science, 32*, 342–365.

Waxman, S. R. (2003). Links between object categorization and naming: Origins and emergence in human infants. In D. H. Rakison & L. M. Oakes (Eds.), *Early category and concept development: Making sense of the blooming, buzzing confusion* (pp. 213–241). London: Oxford University Press.

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