

Linguistic cued attention in children: Words organize attention to shape in a visual search task

Catarina Vales (cvailes@indiana.edu)

Department of Psychological and Brain Sciences, 1101 E 10th St
Bloomington, IN 47405 USA

Linda B. Smith (smith4@indiana.edu)

Department of Psychological and Brain Sciences, 1101 E 10th St
Bloomington, IN 47405 USA

Abstract

By one account of early word learning, children become proficient word learners as a result of environmental regularities: Learning words tunes the child to the regularities offered by the language being learned, orienting attention to those regularities. We test one core claim of this account, that count nouns should cue attention to the shape of the objects. Using a visual search task we present evidence that hearing the name of the object narrows children's attention to the objects in the array that have the same shape. Future steps and the implications of these results are discussed.

Keywords: attention; language and cognition; visual search; word learning.

Introduction

Using past experience to select what to attend to is a powerful feature of human cognition. If exposed to environmental regularities, infants (Kirkham, Slemmer and Johnson, 2002; Saffran, Aslin & Newport, 1996) and adults (Chun & Jiang, 1998; Zhao, Ngo, McKendrick, & Turk-Browne, 2011) readily attend to the current events that better match the underlying structure of their previous experience.

This ability to selectively attend to the most reliable sources of information as a result of past experience has been proposed as a mechanism underlying early word learning. Not only do children learn a large amount of words in the first years of life, but they also seem to do so in very smart ways. For example, by 2.5 years children use shape to generalize new noun categories – if given a novel named object, children will selectively attend to shape over color or texture when extending the novel name to new exemplars (the *shape bias*; Booth & Waxman, 2002; Jones & Smith, 2002).

Because in English many count nouns map to object categories well organized by within-category shape similarity, learning individual word-object mappings could create a top-down process that would organize future learning. According to the attentional learning account of the shape bias, it is the co-occurrence of nouns and shapes that creates an attentional bias to shape over other features when generalizing a new object category. Although there is evidence supporting the attentional account (Gershkoff-Stowe & Smith, 2004; Jones, 2003; Smith, Jones, Landau, Gershkoff-Stowe, & Samuelson, 2002), the specific

mechanisms through which nouns cue object shape have not been directly tested. In the traditional shape bias task, children are asked to generalize a new object category in an untimed forced-choice procedure. The fact that children tend to select the shape match could be due to attentional processes or to much later decision processes. Indeed, competing theories of the attentional learning account have suggested that the shape bias reflects more conceptual theories about how words refer to objects (Waxman & Gelman, 2009).

The purpose of this paper is to empirically test a core claim of the selective attention account: That words cue children's *attention* to the shape of the objects. To this end, we use a visual search task – a well-documented attentional task in which participants are asked to find a particular object (the target) amidst distractor objects. In the visual search literature with adults (Treisman & Gelade, 1980) and children (Gerhardstein & Rovee-Collier, 2002), when the target and the distractors differ by just one feature, search is almost effortless and does not depend on the number of distractors. When the target and the distractors have overlapping features, finding the target becomes a serial search – and response times depend on the number of distractors. The intercept and the slope of the search function are also indicators of the attentional processes involved. While the slope reflects the per item search time (i.e. how long it takes per item to decide if it is the target), the intercept is thought to reflect pre-search processes, including the representation of the search target in working memory (Vickery, King & Jiang, 2005; Woodman, Vogel and Luck, 2001).

To investigate the role of labels in visual attention we use a visual search task to compare children's performance when they were cued with both the spoken name and a picture of the target versus when they were cued with just a picture of the target.

Experiment 1: Do labels cue attention to shape?

Finding a target requires keeping a representation of the target in working memory. Research with adults has suggested that more robust working memory representations of the target result in overall decrease in search times (i.e. intercept changes). Because visual attention is biased

towards elements in the array that match the contents of working memory, stronger working memory representations would effectively suppress attention to nonmatching elements in the search array (Kristjansson, Wang & Nakayama, 2002; Soto & Humphreys, 2007; Vickery, King & Jiang, 2005), therefore modulating pre-search attentional processes.

Does hearing the name of the target prior to search influence its representation in working memory, and thus search? If hearing the object name results in the enhanced representation of object shape relative to other properties, then the explicit naming of the search target on each trial should effectively narrow search to items in the array with the same shape. This is the hypothesis tested in Experiment 1.

In a conjunctive search task, children were asked to search for a particular colored object (e.g., red bed) in a field of same shape (e.g., green bed) and same color (e.g., red couch) distractors. In the Label condition, children heard the displayed object (but not its color) named (e.g., “bed”) prior to each search trial; in the Silent condition, they just saw the displayed target. If storing the name along with the target object in working memory supports processes that automatically direct attention to same shaped items in the array, then overall search time should decrease in the Label condition as participants would preferentially examine the shape matching objects to find the conjunctive match. That is, by hypothesis, in the Label condition children’s attention might be automatically attracted to the shape matching items, with attention to the non-shape matching items being dampened. If so, this would effectively reduce the search set and should lead to faster overall search times in the Label than in the Silent condition.

Methods

Participants. Thirty-two children ($M=37$ months, range: 31-43 months) were assigned to either the Silent or the Label condition. Ten additional children were recruited but not included in the final sample due to refusal to participate in the study ($N=3$), not finishing the familiarization phase ($N=1$), or selecting a non-target object on most test trials ($N=6$). Children were reported to have no developmental disorders, normal (or corrected to normal) visual acuity and color vision. English was the main language spoken by all families. Parental consent was obtained for all participants in compliance with the IRB of Indiana University.

Stimuli and procedure. Figure 1 shows the experimental set up and the temporal order of events on each trial. The child was seated approximately 35cm from a 17” monitor equipped with a touchscreen (MagicTouch, Keytec, Garland, TX). E-Prime software (PST, Pittsburg, PA) was used to control stimulus presentation and record the latency and the location of each response during the test phase. On each test trial, a “fixation” slide encouraged the child to rest their hands on the table (Figure 1a) before the target object was displayed on the center screen for 1sec (Figure 1b).

The search array (with the target object amid distractor objects) was then displayed and the child asked to find the target picture as fast as possible (Figure 1c). Prior to the test phase, children were familiarized with the search procedure, with holding their hands on the table during fixation, and touching the target.

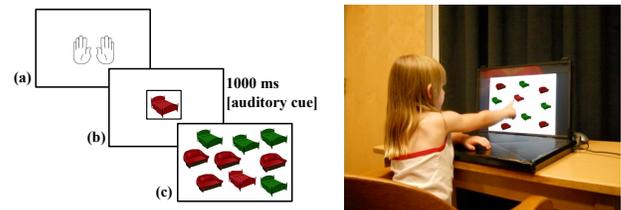


Figure 1: Structure of a trial (left) and child performing the task (right).

Each child was assigned one search target and searched for the same object throughout 32 test trials. Four different objects served between subjects as targets: a red bed, a red couch, a green bed, and a green couch. For each target, the distractors were selected so that half had the same shape and half had the same color as the target (that is, when the target was a red bed, half the distractors were red couches and half were green beds). Each test stimulus was rendered in a 180 x 140 pixel area on a white background. Across trials, the number of distractor objects was manipulated: on each trial, the target object was placed amidst 2, 4, 8 or 12 distractors; eight occurrences of each distractor set size was presented in an order randomly determined for each subject. Sixteen possible locations were used to place the target and the distractors. Across test trials, the target appeared equally often on the left and right side of the screen.

The experimenter started each trial ensuring that the child was looking at the screen; no time limit was set for finding the target. No feedback was given during test phase. In the Label condition, a sound file containing the name of the target object (e.g. “bed”) played at the onset of the target (Figure 1b). The audio files were recorded using an artificial speech creator at a sample rate of 16KHz. No sound file was played in the Silent condition.

Results and Discussion

Mean reaction times (RT) per distractor level were calculated for each child. Only correct responses (i.e. when the target object was selected) were included. Although some participants did not complete all test trials, no differences were found between conditions in the total number of trials completed, $t(30) = -0.37, n.s.$, nor in accuracy, $t(30) = 0.14, n.s.$ (see Table 1). Figure 2 depicts mean RT for the Silent and the Label conditions as a function of number of distractors. A mixed 2 x 4 analysis of variance with condition as the between-subjects factor and number of distractors as the within-subjects factor yielded a main effect of distractor number [$F(3,90) = 27.30, p < 0.001$], reflecting the fact that RT increased as the number

of distractors increased. A main effect of condition was also found [$F(1,30) = 4.48, p < 0.05$], reflecting a significant decrease in overall RT for the Label condition. Number of distractors and condition did not interact [$F(3,90) = 0.21, n.s.$]. The slopes and intercepts of the linear best-fit lines were also calculated for each child. Independent samples t-tests showed that while the slopes of the two conditions were not different [$t(30) = 0.39, n.s.$], there was a significant reduction in the intercept of the Label condition when compared to the Silent condition [$t(30) = -2.40, p < 0.05$].

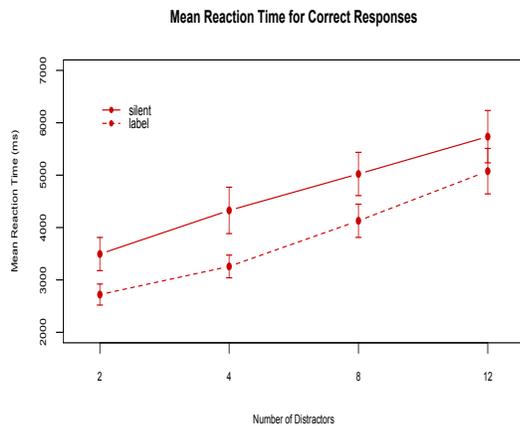


Figure 2: Mean RT (ms) per number of distractors for the Silent and the Label conditions in Experiment 1. Error bars represent standard errors.

These results are consistent with the hypothesis that hearing the name increases attention to shape matching items and/or decreases attention to non-shape matching distractors – thus decreasing overall search time. The results provide direct evidence for a role of object names in guiding children’s attention to object shape. However, presenting the target label did not affect the slope of the search function, which may indicate that the label does not affect the time it takes to make a decision per each attended item. This point will be addressed in the General Discussion section.

Experiment 2 examines an alternative account for the present findings; that it was not the object name per se which enhanced overall search time, but the presence of an auditory signal at the start of each trial.

Experiment 2: Does any word cue attention?

A growing literature shows multimodal influences on visual attention and search such that auditory cues may lead to more rapid search (Iordanescu, Grabowecy, & Suzuki, 2011; Van der Burg, Olivers, Bronkhorst, & Theeuwes 2008). Thus it is possible that the effects observed in Experiment 1 were due to the addition of a spoken word – potentially any word – and not to the target’s name nor increased attention to object shape. Accordingly, Experiment 2 replicated the Label condition of Experiment

1 but replaced the target name on each trial with the word “Go.”

Methods

Participants. Sixteen children between 32 and 42 months of age ($M = 37$ months) participated; none of these children had participated in Experiment 1. Eleven additional children were recruited but not included in the final sample due to selecting a non-target object on most test trials. Recruitment and informed consent procedures were the same as in Experiment 1.

Stimuli and procedure. All aspects were the same as in the Label condition of Experiment 1, except that the sound file presented at the onset of the target played the word “Go.”

Results and Discussion

Mean RT per number of distractors for correct responses was calculated for each child. Children completed 30 trials ($SD = 2.98$) on average, and mean accuracy was 83% (see Table 1). Figure 3 presents RT for correct responses per distractor level for the Go condition. For comparison purposes, results from the Silent condition from Experiment 1 are also shown. A mixed 2×4 analysis of variance with number of distractors as within-subjects factor and condition as the between-subjects factor yielded no reliable differences in RT between the Go condition of Experiment 2 and the Silent condition of Experiment 1 [$F(1, 30) = 0.06, p = 0.82$]. A main effect of distractors number was found [$F(3, 90) = 23.82, p < 0.001$], reflecting the increase in RT as a result of increasing the number of distractors. These two factors did not interact [$F(3, 90) = 0.41, p = 0.75$]. The analyses of the individual slopes and the intercepts confirmed the trends found for RT: No differences were found between the Go condition of Experiment 2 and the Silent condition of Experiment 1 in the slope [$t(30) = 0.25, p = 0.80$] or the intercept [$t(30) = -0.38, p = 0.71$].

In brief, an auditory word that is not the name of the target does not result in more rapid search than the presentation of no sound at all, a result that suggests the observed effects in Experiment 1 were not due to an auditory cuing effect but instead reflected the presentation of the object name.

Table 1: Experiments 1 and 2 – Slopes and Intercepts of the search functions, Mean accuracy and Mean number of trials completed.

Condition	Slope (SE)	Intercept (SE)	Accuracy (SE)	Trials completed (SD)
Exp.1-Silent	212 (28)	3264 (212)	85 (3)	31 (1.55)
Exp.1-Label	233 (5)	2284 (37)	86 (4)	31 (3.75)
Exp.2-Go	223 (15)	3085 (113)	83 (3)	30 (2.98)

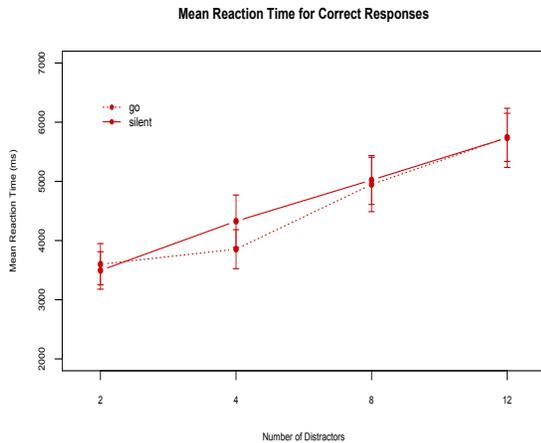


Figure 3: Mean RT (ms) per number of distractors for the Go condition of Experiment 2 and the Label condition in Experiment 1. Error bars represent standard errors.

General Discussion

Word learning requires selective attention to the right properties – shape is likely to be the right feature if generalizing countable objects, while texture might be a better alternative if generalizing food terms. What processes support the development of these attentional biases? This question is the source of a major dispute in the literature, with some arguing for the role of attended regularities in setting up such attentional biases (Smith, Jones, Yoshida, Colunga, 2003) and others arguing that word learning entails more conceptual and deliberative processes about how categories are formed (Waxman & Gelman, 2009).

The current results offer support to the first alternative. By showing that words direct *attention* to the shape of known objects, we offer for the first time direct evidence for a core claim of the attentional learning account. After enough instances of attention being directed to the shape of known objects in the presence of their names, a generalized attentional bias might emerge – any noun, even a novel one, could now cue attention to shape. This is potentially a very powerful learning mechanism, one that enables children to quickly generalize new categories in the presence of new words.

The current results also suggest that the attentional effects of words may be located at the level of working memory – hearing an object’s name strengthens the shape representation of that object. The finding that hearing the target name on a visual search task influences the intercept of the search function, without changes in the slope, also suggests that the effects of words do not influence the efficiency of search (i.e. how long it took to identify or dismiss each item as the target). However, the present version of the task may not be optimal to test for a potential role of labels in the efficiency of search (as measured by the slope of the search function). In adults, the slope – or per item search time – is affected by the familiarity of the target and distractors (Mruczek & Sheinberg, 2005; Wang,

Cavanagh & Green, 1994), by the need or ease of binding the features of individual items in the array into their individual units (Treisman & Gelade, 1980), and by the discriminability of the target from the distractors (Duncan & Humphreys, 1989). For young children, providing the basic-level name of the target could, in principle, influence any of these processes – and in so doing increase the role of shape in search in ways that expedite the identification of the target. By this hypothesis, given sufficiently difficult shape discriminations, hearing the object name prior to search might be expected to yield a decreased slope in the search function as well as a decrease in overall search time. This is a critical issue for future research.

What are the implications of the current results? Although more research is needed to further understand the mechanisms involved in the attentional effects of labels in word learning, the evidence presented here suggests that hearing a name activates a representation of certain features of the object – in the case of count nouns, object shape. It follows that hearing an object name will cue attention to that object’s shape, and over time this has the potential to not only become a more automatic process, but also to change the nature of the representation (possibly from specific individual features to more abstract shape representations). Moreover, by extension from accounts of these processes in the adult literature (Dahan & Tanenhaus, 2005; Huettig & Altmann, 2007; Lupyan & Spivey, 2010) these labeling effects appear to be rapid and automatic, that is, not under deliberative or conceptual control. Thus, the current results provide a stepping-stone to a mechanistic account of how words organize attention in children – and in so doing, may organize early word learning and the on-line comprehension of words in context.

Acknowledgments

This research was supported by a grant from the National Institute of Child Health and Development (HD28675) to LBS and a Graduate Fellowship from the Portuguese Foundation for Science and Technology awarded to CV (SFRH/BD/68553/2010). The authors would like to thank the members of the Cognitive Development Lab at IU for useful discussions, Anna MacKinnon, Blakely Meyer and Tracy Kelsey for their help with recruitment and data collection, and the parents and children who participated in these studies.

References

- Booth, A. E., & Waxman, S. R. (2002). Word learning is ‘smart’: evidence that conceptual information affects preschoolers’ extension of novel words. *Cognition*, 84(1), B11–B22.
- Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology*, 36, 28–71.
- Dahan, D., & Tanenhaus, M. K. (2005). Looking at the rope when looking for the snake: Conceptually mediated eye

- movements during spoken-word recognition. *Psychonomic Bulletin & Review*, 12(3), 453–459.
- Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, 96(3), 433.
- Gerhardstein, P., & Rovee-Collier, C. (2002). The development of visual search in infants and very young children. *Journal of Experimental Child Psychology*, 81: 194–215.
- Gershkoff-Stowe, L., & Smith, L. B. (2004). Shape and the first hundred nouns. *Child Development*, 75(4), 1098–1114.
- Huettig, F., & Altmann, G. T. M. (2007). Visual-shape competition during language-mediated attention is based on lexical input and not modulated by contextual appropriateness. *Visual Cognition*, 15(8), 985–1018.
- Iordanescu, L., Grabowecky, M., & Suzuki, S. (2011). Object-based auditory facilitation of visual search for pictures and words with frequent and rare targets. *Acta Psychologica*, 137(2), 252–259.
- Jones, S. S. (2003). Late talkers show no shape bias in a novel name extension task. *Developmental Science*, 6(5), 477–483.
- Jones, S. S., & Smith, L. B. (2002). How children know the relevant properties for generalizing object names. *Developmental Science*, 5(2), 219–232.
- Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: evidence for a domain general learning mechanism. *Cognition*, 83(2), B35–B42.
- Kristjansson, A., Wang, D. & Nakayama, K. (2002). The role of priming in conjunctive visual search. *Cognition*, 85, 37–52.
- Lupyan, G., & Spivey, M. J. (2010). Redundant spoken labels facilitate perception of multiple items. *Attention, Perception & Psychophysics*, 72(8), 2236–2253.
- Mruczek, R. E. B. & Sheinberg, D.L. (2005). Distractor familiarity leads to more efficient visual search for complex stimuli. *Perception and Psychophysics*, 67(6), 1016–1031.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274(5294), 1926–1928.
- Smith, L. B., Jones, S. S., Landau, B., Gershkoff-Stowe, L., & Samuelson, L. (2002). Object name Learning Provides On-the-Job Training for Attention. *Psychological Science*, 13(1), 13–19.
- Smith, L.B., Jones, S.S., Yoshida, H., Colunga, E (2003). Whose DAM account? Attentional learning explains Booth and Waxman. *Cognition*, 3, 209–13.
- Soto, D. & Humphreys, G. W. (2007). Automatic guidance of visual attention from verbal working memory. *Journal of Experimental Psychology: Human Perception and Performance*, 33(3), 730–753
- Treisman, A.M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12: 97–136.
- Van der Burg, E., Olivers, C. N. L., Bronkhorst, A. W., & Theeuwes, J. (2008). Pip and pop: Nonspatial auditory signals improve spatial visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 34(5), 1053–1065.
- Vickery, T. J.; King, L.; Jiang, Y. (2005). Setting up the target in visual search. *Journal of Vision*, 5, 81–92.
- Wang, Q., Cavanagh, P., Green, M. (1994). Familiarity and pop-out in visual search. *Perception and Psychophysics*, 56(5), 495–500.
- Waxman, S. R., & Gelman, S. A. (2009). Early word-learning entails reference, not merely associations. *Trends in Cognitive Sciences*, 13(6), 258–263
- Woodman, G.F.; Vogel, E.K.; Luck, S.J. (2001). Visual Search Remains Efficient When Visual Working Memory Is Full. *Psychological Science*, 12(3), 219–224.
- Zhao, J., Ngo, N., McKendrick, R., & Turk-Browne, N. B. (2011). Mutual interference between statistical summary perception and statistical learning. *Psychological Science*, 22, 1212–1219.