



ELSEVIER

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

journal homepage: www.elsevier.com/locate/jecp



When a word is worth more than a picture: Words lower the threshold for object identification in 3-year-old children



Catarina Vales*, Linda B. Smith

Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN 47405, USA

ARTICLE INFO

Article history:

Received 25 September 2017

Revised 30 March 2018

Available online 5 July 2018

Keywords:

Language

Vision

Object recognition

Visual perception

Language development

Cognitive development

ABSTRACT

A large literature shows strong developmental links between early language abilities and later cognitive abilities. We present evidence for one pathway by which language may influence cognition and development: by influencing how visual information is momentarily processed. Children were asked to identify a target in clutter and either saw a visual preview of the target or heard the basic-level name of the target. We hypothesized that the name of the target should activate category-relevant information and, thus, facilitate more rapid detection of the target amid distractors. Children who heard the name of the target before search were more likely to correctly identify the target at faster speeds of response, a result that supports the idea that words lower the threshold for target identification. This finding has significant implication for understanding the source of vocabulary-mediated individual differences in cognitive achievement and, more generally, for the relation between language and thought.

© 2018 Elsevier Inc. All rights reserved.

Introduction

Many traditional theories of cognitive development regard language as a transformative force on cognition, enabling children to think abstractly, to control their cognitive processes, and to reason about instances that cannot be directly perceived. As [Vygotsky \(1980\)](#) famously noted, “The child

* Corresponding author at: Department of Psychology, Carnegie Mellon University, Pittsburgh, PA 15213, USA.

E-mail address: cvalues@andrew.cmu.edu (C. Vales).

begins to perceive the world not only through his eyes but also through his speech. As a result, the immediacy of ‘natural’ perception is supplanted by a complex mediated process.” Here we show evidence consistent with the idea that language changes other cognitive abilities, and we provide one mechanism by which words may do so in children as young as 3 years — by changing in-the-moment processing of visual information.

A large literature shows that children’s language proficiency is related to their cognitive skills. Early individual differences in vocabulary size and semantic knowledge predict *later* intelligence (Marchman & Fernald, 2008), school readiness (Morgan, Farkas, Hillemeier, Hammer, & Maczuga, 2015), and behavior regulation (Petersen et al., 2013). Although these relations have been established, the developmental pathways through which language influences other cognitive abilities are not known. In the laboratory, heard words have been shown to influence how long children sustain attention on an object (Baldwin & Markman, 1989), how children compare and categorize visual objects (Ferry, Hespos, & Waxman, 2010; Yoshida & Smith, 2003), and how children reason about relations within and among objects (Dessalegn & Landau, 2013; Loewenstein & Gentner, 2005). Although these results suggest that language influences foundational cognitive processes, they do not elucidate the specific mechanisms by which language learning affects so many aspects of cognition.

One pathway through which language and language learning could influence performance in different domains is by influencing how visual information is momentarily processed. This idea has been considered in the context of traditional debates on whether language can influence perception or whether human cognition is based on nonmalleable percepts (Gleitman & Papafragou, 2012). Although we consider these classic questions in the Discussion section, our hypothesis about how language influences the online processing of visual information is based on findings from adults suggesting that words have direct effects on visual processes (Lupyan & Spivey, 2010; Ostarek & Huettig, 2017). Most human learning and performance has substantial visual components (Ahissar & Hochstein, 2004). Thus, if words can influence how children momentarily process visual information, this would constitute a plausible mechanism for how language could have pervasive effects on other cognitive abilities. We investigated this question by examining how children identify a target object in clutter, one domain in which words have been shown to affect visual processing in adults.

When presented with a cluttered array of objects and the goal of finding a particular target, the perceiver needs to compare the incoming visual information and the array with a memory or representation of the target (Hout & Goldinger, 2015). Adult performance in these tasks depends on the information provided *before* the search array is presented (Vickery, King, & Jiang, 2005), with naming the target being positively related to target identification (Lupyan & Spivey, 2010). This is thought to reflect a process whereby words activate visual memories about the categories of objects to which they refer (Lupyan, 2008; Lupyan & Thompson-Schill, 2012; see also Jonides & Gleitman, 1972), including relevant low-level visual information that lowers the threshold to identify an object (Ostarek & Huettig, 2017). Here we asked whether words have similar effects on 3-year-old children’s identification of objects in clutter. On the one hand, these young children are at the beginning of the long developmental trajectory of word learning and visual object recognition (Smith, 2003) and differ from older children and adults in how they process visual objects (e.g., Mash, 2006). Thus, it is possible that words will not influence how such young children visually identify objects. The alternative possibility, that words do influence how children identify objects in clutter, would constitute a potential causal explanation for the power of language in cognitive development.

In the current experiment, 3-year-old children searched for a target that was an instance of a category (e.g., ice cream) among distractors that were instances of other visually similar categories (e.g., balloons, lamps). In the Label condition, children heard the spoken name of the target object prior to search; in the Visual Preview condition, children saw a visual preview of the target object prior to search. The experimental question was this: When identifying a target in an array of visually similar distractors, does hearing the basic-level name of the target object lower the threshold to identify the target *relative to* seeing a visual preview of the actual target? Both hearing the name and seeing a preview of the target have been shown to facilitate target identification in adults (Lupyan & Spivey, 2010; Vickery et al., 2005), and the combination of a label and a visual preview has been shown to increase the speed with which 3-year-old children identified a target object relative to a visual preview (Vales & Smith, 2015). Here we tested the more consequential prediction that the name of

the target alone facilitates target identification more than a visual preview of the to-be-found object. Our reasoning is this: If the name of the target activates category-relevant information for object identification, then hearing the name of the target object should activate a better visual representation of the target than seeing a visual preview of the target. This is because the visual preview of the target will activate both category-general and item-specific information (e.g., [Hollingworth, Williams, & Henderson, 2001](#)); this item-specific information, by not discriminating target and distractors, should make target identification more difficult.

To test the hypothesis that labels lower the threshold for target identification, we used a child-friendly version of a response deadline task. This approach, by forcing participants to respond rapidly, measures the minimal time needed to make a correct decision given the presented information (e.g., [Pachella, Fisher, & Karsh, 1968](#)). If the threshold for identifying the target is lowered by hearing the name of the target, then the minimal time needed to identify the target should be less in the Label condition than in the Visual Preview condition. The key prediction, thus, concerns rapid responding; when responding *rapidly*, children should be more likely to respond *correctly* in the Label condition than in the Visual Preview condition. While fixed brief deadlines and the possibility of many errors are typical of deadline tasks, these task properties are unlikely to yield useful data from 3-year-old children. Accordingly, instead of imposing deadlines, we encouraged children to respond rapidly but let children self-pace their responses. Similar to the analytic approach of deadline tasks, we then systematically related the time taken to respond to the likelihood of correctly identifying the target across the two conditions. Although this approach does not manipulate response deadlines, it nevertheless allows us to examine the relation between time taken to respond and accuracy, a relation that is informative of how children identify a target in clutter when shown its visual preview versus when hearing its name (see [Fitts, 1966](#); [Ratcliff, 1985](#)), and whether the minimal time needed to make a correct decision is lowered by hearing the name of the target ([Pachella et al., 1968](#)).

Method

Participants

Fifty-two children between 32 and 41 months of age (23 girls; $M = 37$ months, $SD = 2.4$), were randomly assigned to the Label condition or the Visual Preview condition, and to one of two targets (ice cream or cup), with an equal number of children assigned to each condition and target. Calculating power for the intended generalized linear mixed-effects analyses relating response time (RT) to accuracy is not trivial ([Johnson, Barry, Ferguson, & Müller, 2015](#)). In addition, to the best of our knowledge, there are no prior studies with child participants relating RT to the likelihood of making a correct response to provide an expected effect size. Accordingly, we approximated our sample size by calculating the target sample size for a traditional linear regression model assuming a small effect size (.15), an alpha level of .05, and a power level of .80; this yielded a total sample size of 55. To satisfy the between-participants assignment to condition and target, we recruited 52 children. Children were recruited from a predominantly working- and middle-class population in Bloomington, Indiana in the midwestern United States and were tested in the laboratory. English was the main language spoken by all families. An additional 13 children were recruited but not included in the analyses due to parental interference ($n = 2$), refusal to participate or failure to complete all test trials ($n = 10$), or English not being the child's main language ($n = 1$). Parental consent was obtained for all children in compliance with the institutional review board of Indiana University, and all children received a toy for participating. Parents were asked to complete the CDI-III ([Fenson et al., 2007](#)), a normed productive vocabulary checklist with 100 words; six families did not complete or return the inventory. Children in the two conditions did not differ in the mean number of words in their productive vocabularies ($M_{\text{Visual}} = 71$, $M_{\text{Label}} = 70$), $t(44) = 0.23$, $p = .81$, Cohen's $d = 0.07$.

Stimuli and design

Two target categories were used: ice cream and cup. For each target, the distractors were items of other categories. For the target ice cream, the distractors were tokens of lamps, balloons, and glasses;

for the target cup, the distractors were tokens of hats, pots, and balls. To test the hypothesis that hearing the name of the target object, by not including item-specific information, activates a better representation of the target than seeing its visual preview, the distractor categories were selected to be visually hard to discriminate from the target while sharing minimal phonological overlap with the name of the target (see Vales & Smith, 2015, Experiment 3). All categories of items used are highly familiar to children of this age (see Vales & Smith, 2015, Experiment 3), and with one exception (“pot”) their names are known by at least 50% of children before 30 months of age (Fenson et al., 2007). To prevent color information from guiding search, the pictures of targets and distractors were recolored in red scale; pictures were rendered in a 100×90 -pixel area on a white background. The audio files used to present the name of the target in the Label condition were recorded using an artificial speech creator at a sample rate of 16 kHz.

Children completed 36 test trials in which they searched for the same target category. The target was always present, and across test trials it was displayed equally often on the left and right sides of the screen. To maintain the difficulty of the task and increase the variability of the RT measure, the number of distractor items varied across trials and multiple tokens of the target and distractor categories were used. Within each block of 9 trials, there were three occurrences of each distractor set size (3, 9, or 12 distractors), with the order being randomized for each child. On each trial, there was the same number of tokens of each distractor category (e.g., for set size 9, there were three tokens of each of the three distractor categories). For each target and distractor category, 18 unique tokens were used (see Fig. 1 in the online supplementary material). Within each block of 18 trials, there was one occurrence of each target/distractors token pairing.

Procedure

Children sat in front of a 17-inch monitor equipped with a touchscreen layover (MagicTouch, Keytec, Garland, TX, USA). E-Prime (Psychology Software Tools, Pittsburgh, PA, USA) was used to present the stimuli and to record participants' responses. Fig. 1 depicts the structure of a single trial. On each trial, a “fixation” slide was presented to prompt children to rest their hands on the table below the screen; this ensured that children's manual responses had similar starting points across trials. The experimenter pressed a keyboard key to present the cue screen once children were looking at the screen; participants in the Label condition were cued with the spoken name of the target object (e.g., “ice cream”) while a blank screen was on, and participants in the Visual Preview condition were cued with a silent visual preview of the exact target object for that trial (see Fig. 1 in the supplementary material). The search array was displayed after 1 s, and children were asked to identify the target picture and touch it; the trial ended once a manual response was detected. No time limit was imposed, but children were reminded to find the target picture as quickly as possible. In giving task instructions, the experimenter did not label any of the objects, instead saying “Watch the picture!” or “Listen up!” before the cue slide and “Can you find it?” during the response screen. No feedback was given. Children were given neutral encouragement regardless of performance (e.g., “Thanks for your help finding the pictures”) and were given stickers to maintain their interest in the task; the stickers were put inside a container and kept out of children's view. Before the experimental trials, children completed a familiarization phase to be acquainted with preparing for the upcoming trial by keeping their hands on the table, listening to the audio cue (in the Label condition) or watching the visual cue (in the Visual Preview condition) before the search array was displayed, and touching the target as soon as they identified it on the search screen; during this familiarization phase, children were given feedback as needed. Pictures of smiley faces, crayons, pencils, bicycles, and motorcycles were used as targets and distractors during this familiarization phase.

Results

Because we were interested in examining the relation between RT and accuracy across the two experimental conditions, we first examined whether the two conditions differed on each individual measure. We started with an analysis of children's RT in the two conditions, regardless of whether

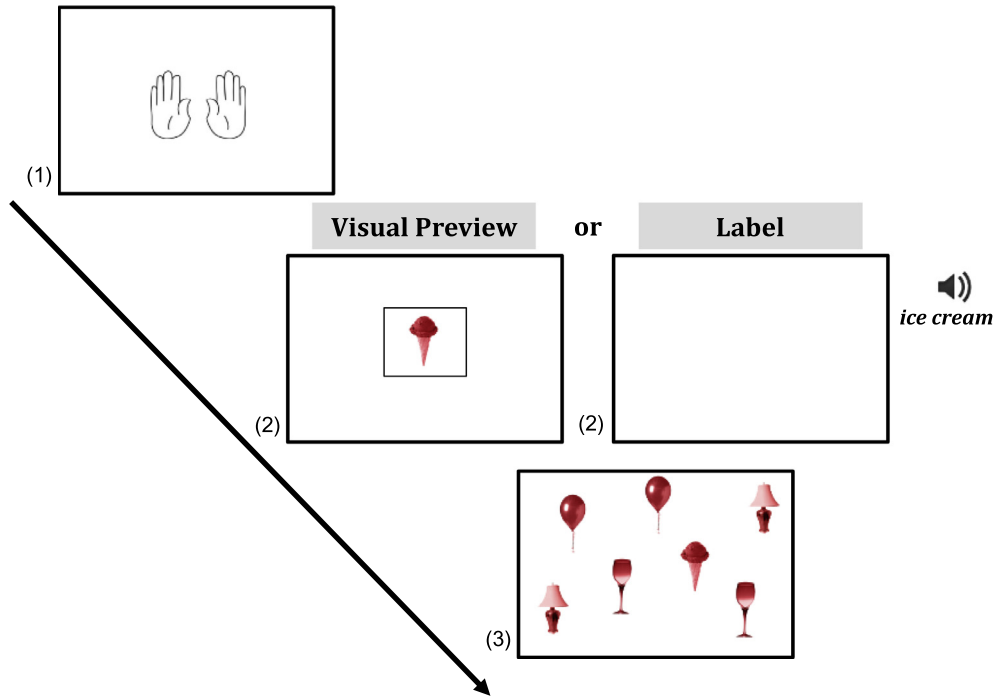


Fig. 1. Structure of a trial in the experiment. Participants in the Visual Preview condition were presented with the visual preview in silence, and participants in the Label condition were presented with the spoken name of the target while a blank screen was on. After 1 s, participants were presented with the search array and were asked to identify the target by touching it.

the RTs were associated with correct or incorrect responses, followed by an analysis of children's accuracy in the two conditions irrespective of time taken to respond. These analyses showed that, when considering RT and accuracy separately, the two conditions did not differ in how long children took to make a response or in how likely children were to correctly identify the target object. We then tested our hypothesis by examining the relation between RT and accuracy; this analysis showed that time taken to respond differentially affected how accurate participants were across the two conditions.

Response time measures

In the adult literature, RT is used to index processing time for correct responses; the longer it takes the perceiver to respond, the more effortful it is to process the presented information for the task at hand (Pachella, 1974). Within the current approach of measuring accuracy as a function of RT, we need RT to reflect the processing time required to yield a correct response; however, by not imposing response deadlines, there could be trials in which children took a long time to respond that did not reflect effortful processing but rather reflected being off-task. However, because of their skewed distributions, it is not trivial to identify outliers in RT data (Ratcliff, 1993; Whelan, 2008), and this may be particularly problematic when a small number of observations is collected from each participant (Ratcliff, 1993), as is often the case with child participants. Given these considerations, we excluded only trials in which children took longer than 10 s to make a response; this threshold was decided on visual inspection of the RT distribution (see Figs. 2 and 3 in the supplementary material) and resulted in the exclusion of 5.6% of all data. We also excluded trials in which children took less than 500 ms to respond, resulting in the additional exclusion of 0.96% of all data (see Table 1 in the supplementary material for number of trials included per condition). These exclusion criteria are similar to

those used in past work with young children (e.g., Frank, Sugarman, Horowitz, Lewis, & Yurovsky, 2016), and (as described below) all results remained unchanged when we included these trials.

Table 1 shows descriptive statistics for each condition, and Fig. 2 depicts the RT distributions for all trials (correct and incorrect) across the two experimental conditions after exclusions. The two conditions did not differ in how long participants took to make a response, $t(50) = 0.30$, $p = .77$, Cohen's $d = 0.08$. To further examine potential differences between the two conditions, their distributions were compared using a bootstrapping version of the Kolmogorov–Smirnov test with 1000 samples; this test showed that the RT distributions of the two conditions were not significantly different ($D = 0.028$, $p = .89$). We also examined whether the number of distractor items across trials differentially influenced how long children took to make a response (regardless of accuracy) across the two conditions (see Table 1). An analysis of variance (ANOVA) with distractor set size (3, 9, or 12) as a within-participant factor and condition (Label or Visual Preview) as a between-participants factor revealed a main effect of distractor set size on RT, $F(2, 100) = 69.3$, $p < .001$, $\eta^2 = .22$, but no interaction between distractor set size and condition, $F(2, 100) < 1$, $p = .88$, $\eta^2 < .001$. Both analyses produced equivalent results when all trials were included (see Additional Analyses A in the supplementary material). Together, these analyses suggest that there were no overall RT differences between the two conditions and that the number of items in the array did not differentially affect how long children took to make a response across the two conditions.

Accuracy measures

The mean proportion of correct responses was calculated for each participant by dividing the number of trials in which the target was correctly identified by the total number of trials in the task. As can be seen in Table 1, children were overall accurate at identifying the target, which indicates an

Table 1
Response time and accuracy measures in the two experimental conditions.

		Label condition		Visual preview condition		
		<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	
All trials	Overall response time (ms)	4358	259	4430	272	
	Response time per distractor set size (ms)	3 distractors	3495	205	3586	236
		9 distractors	4466	312	4529	292
		12 distractors	5113	325	5175	331
	Overall proportion of correct responses	.90	.02	.87	.03	
	Proportion of correct responses per distractor set size	3 distractors	.92	.03	.91	.02
		9 distractors	.90	.02	.85	.03
		12 distractors	.90	.03	.84	.04
After RT outlier exclusion	Overall response time (ms)	3884	169	3954	161	
	Response time per distractor set size (ms)	3 distractors	3286	165	3306	148
		9 distractors	3982	202	4043	207
		12 distractors	4430	204	4551	189
	Overall proportion of correct responses	.91	.02	.89	.03	
	Proportion of correct responses per distractor set size	3 distractors	.92	.02	.92	.02
		9 distractors	.91	.02	.88	.03
		12 distractors	.90	.02	.87	.04

Note. The top section displays the summary for data from all trials, and the bottom section displays the summary data for data without trials in which children took less than 500 ms or more than 10 s to make a response.

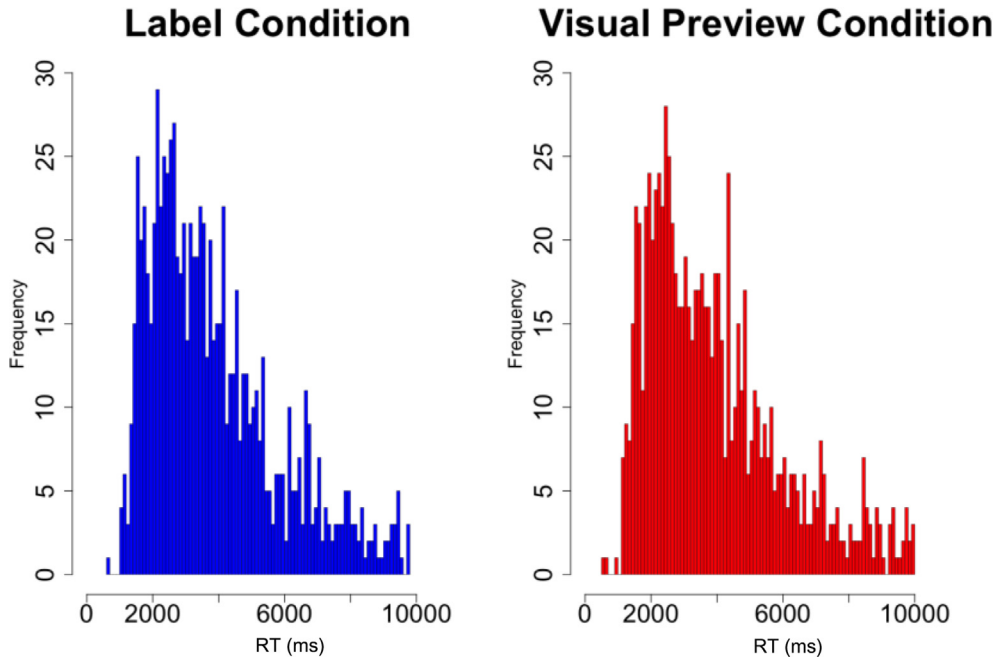


Fig. 2. Response time distributions (ms) for the two experimental conditions for both correct and incorrect responses.

understanding and commitment to the task. The two conditions did not differ in the overall likelihood of identifying the target, $t(50) = -1.02$, $p = .31$, Cohen's $d = -0.29$. To assess whether the number of distractors influenced the two conditions differently, an ANOVA with set size (3, 9, or 12) as a within-participant factor and condition (Label or Visual Preview) as a between-participants factor revealed a main effect of distractor set size on accuracy, $F(2, 100) = 4.90$, $p = .009$, $\eta^2 = .02$, but no interaction between set size and condition, $F(2, 100) = 1.15$, $p = .32$, $\eta^2 = .005$ (see Table 1). Both analyses produced equivalent results when we included only trials in which children took more than 500 ms or less than 10 s to respond (see Additional Analyses B in the supplementary material). Together, these analyses show that children in both conditions identified the target in clutter at similar rates across the two conditions and that the number of items in the visual array affected children's ability to identify the target to a similar extent in both conditions.

Relation between response time and accuracy

The key prediction concerns how accuracy relates to RT in the two conditions. If words influence visual processing by changing the threshold to accurately identify the target, then any facilitation in finding the target after hearing its name relative to seeing its visual preview should be more pronounced at faster speeds of response. Fig. 3 depicts the relations between time taken to make a response and likelihood of correctly identifying the target; for the purpose of displaying the data, the RT variable was aggregated into three bins with approximately the same number of trials each (see Table 1 in the supplementary material). Children in the Visual Preview condition were more likely to correctly identify the target when they took longer to make a response, whereas children in the Label condition were more accurate at faster speeds of response; in addition, children in the Label condition also seemed more likely to be correct at the fastest RTs than children in the Visual Preview condition. To analyze these patterns, we used a generalized linear model approach (Liang & Zeger, 1986; Zeger & Liang, 1986) to model the relation between RT and accuracy across the two experimental conditions; this allowed us to model the RT measure as a continuous predictor of a repeated

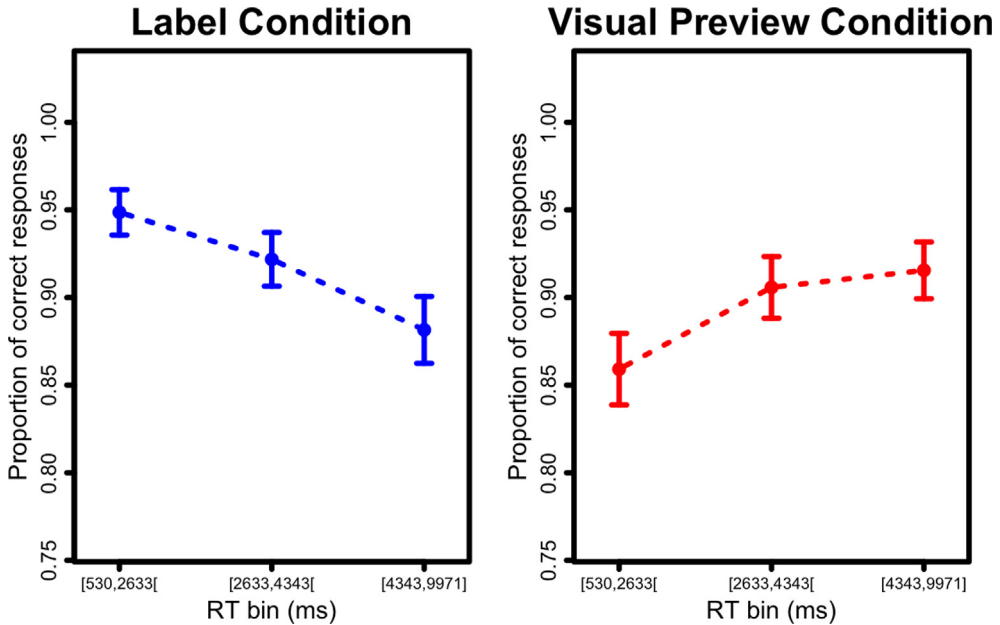


Fig. 3. Mean proportion of correct responses per response time bins in the two conditions. Error bars display standard errors of the mean. The RT bins were calculated for a single distribution that included the two conditions. The RT data were only binned to graphically display the relation between the two variables; the RT per trial was entered into the model as a continuous variable.

categorical response (i.e., for each participant, correct or incorrect response at the trial level) and to examine whether their relation was different for the two conditions. We implemented the model in the R environment (R Core Team, 2014) using the `geeglm` function from the `geepack` package (Højsgaard, Halekoh, & Yan, 2006). To model the binary outcome (i.e., correct or incorrect response), we used a logit link function and a binomial variance function. Because we did not have a priori hypotheses regarding how responses were correlated across trials, we used an independent correlation structure (see also Zeger & Liang, 1986, for how generalized linear models are robust to misspecifications of the correlation structure). The variables RT (as a continuous variable) and condition (Label or Visual Preview) were sequentially entered, followed by an interaction term; the RT variable was centered to decrease the differences in the scales of the model parameters. The `anova` function was used to calculate sequential Wald tests and respective p values for each added variable. Both RT [$b = -0.0002$, $STE = 0.00005$, Wald $\chi^2(1) = 1.68$, $p = .19$] and condition [$b = -0.4$, $STE = 0.4$, Wald $\chi^2(1) = 0.61$,

$p = .44$] did not have an effect on accuracy. However, the interaction between RT and condition had an effect on accuracy [$b = 0.0003$, $STE = 0.00009$, Wald $\chi^2(1) = 12.13$, $p = .0005$], showing that the relation between the time taken to respond and the likelihood of correctly finding the target was not the same across the two experimental conditions. Adding distractor set size to the model did not change the overall results (see Additional Analyses C in the supplementary material). Finally, the interaction between RT and condition still had a significant effect on accuracy when the model included all trials (i.e., when it included trials taking less than 500 ms and more than 10 s; see Additional Analyses D in the supplementary material).

Taken together, the results are consistent with the hypothesis that words lower the threshold to correctly identify an object. These results show that although the two conditions do not differ overall in how long children take to make a response or how likely children are to find the target, the two conditions do differ in the function that relates accuracy to response times. Participants in the Label

condition were very accurate at finding the target at very fast RTs, whereas children in the Visual Preview condition were more accurate when they took longer to make a response.

Discussion

The results show that words influence the speed with which children can identify a target in a cluttered array, with children being more likely to correctly identify the target at faster speeds of response when a word, relative to a visual preview, was presented; this concurs with the hypothesis that words lower the threshold to identify a visual object. The clear implication is that words create an internal target representation that is in some way different from the representation created by seeing the actual target during the visual preview. Lastly, the results support the more specific idea that, relative to visual information, words may be better activators of category-defining properties that distinguish the target from items of other categories. These findings have important consequences for understanding the role of language in learning and development and for understanding individual differences in cognitive development.

Although prior works show that words change children's performance in many laboratory tasks (Baldwin & Markman, 1989; Loewenstein & Gentner, 2005; Yoshida & Smith, 2003), these do not locate the effects at the level of object identification. Motivated by prior findings from the adult literature, we hypothesized that by activating category-relevant information, words would lower the threshold to identify an object in clutter. Contemporary research shows the reach of top-down information in influencing early visual areas (Ahissar & Hochstein, 2004; Bar et al., 2006; Rahman & Sommer, 2008), including cascading effects into areas that support discrimination along category-relevant dimensions (De Baene, Ons, Wagemans, & Vogels, 2008; Folstein, Palmeri, & Gauthier, 2012). The demonstration that these effects hold even in very young children who are still building vocabularies and category knowledge (Smith, 2003) is a significant new contribution because these effects of words on momentary visual object detection have the potential to fundamentally change what children learn and how they perform across many different cognitive tasks.

Words that change the speed with which objects are detected in clutter will have powerful effects, influencing how well children can link the words they hear to specific referents and elements in the visual world, and supporting online sentence comprehension and learning in informal learning settings and in school (Ferreira, Apel, & Henderson, 2008; Marchman & Fernald, 2008). Indeed, this one effect may be the ultimate explanation of why words influence how children remember and generalize newly learned object names (Samuelson, 2002; Yoshida & Smith, 2005), the rate at which they learn new words (Perry, Samuelson, Malloy, & Schiffer, 2010), and the specific words they learn next (Colunga & Sims, 2017) – effects that likely underlie the predictive relations between early vocabulary size and later cognitive achievements. If a heard known word momentarily changes visual processing with effects on what is detected in a cluttered world and later remembered, then these individual moments of experience may accumulate to yield a visual system with stronger top-down effects – and, thus, not just quicker visual processing in the moment but also fundamentally different visual inputs to other cognitive systems (Oakes, 2017). Although these effects may be small in any one moment, the aggregate of many small benefits of words constitutes a plausible mechanism for how early language proficiency is such a strong predictor of later cognitive achievements.

The current results are also relevant for the larger theoretical discussion on language and thought and whether language shapes human cognition or is impermeable to linguistic influences. Traditionally, this question has been investigated by comparing adult speakers of different languages to examine whether they perform differently in nonverbal tasks that require distinctions more strongly highlighted by one of the languages (Gleitman & Papafragou, 2012). If they do, then differences in the task must result from experience with different languages, and thus language can change cognition. However, disentangling and reconciling the contributions of linguistic and nonlinguistic processes is not trivial (Li & Gleitman, 2002; Pederson et al., 1998). Although we did not show that language changes performance in a purely nonlinguistic task – the gold standard within some theoretical frameworks for showing an effect of language on perception – the results did show that the category name speeds the detection of a visual target relative to seeing the specific target to be found. If words systematically influence the momentary processing of information even in just this one way,

in the long term they will alter what children learn and know about the world. Because different languages differ in the categories they lexicalize, the accumulation of experiences with words and scenes could yield differences in object recognition (Kuwabara & Smith, 2016), as well as in more general knowledge.

Acknowledgments

The authors thank Elizabeth Clerkin, Afiah Hasnie, and Adriana Valtierra for their help with this project, the parents and children who participated in this study, and the Indiana University Statistical Consulting Center. This work was supported by a grant from the National Institute of Child Health and Development to LBS and a graduate fellowship from the Portuguese Science Foundation to CV.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.jecp.2018.04.015>.

References

- Ahissar, M., & Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends in Cognitive Sciences*, 8, 457–464.
- Baldwin, D. A., & Markman, E. M. (1989). Establishing word–object relations: A first step. *Child Development*, 60, 381–398.
- Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, A. M., ... Halgren, E. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences of the United States of America*, 103, 449–454.
- Colunga, E., & Sims, C. E. (2017). Not only size matters: Early-talker and late-talker vocabularies support different word-learning biases in babies and networks. *Cognitive Science*, 41, 73–95.
- De Baene, W., Ons, B., Wagemans, J., & Vogels, R. (2008). Effects of category learning on the stimulus selectivity of macaque inferior temporal neurons. *Learning & Memory*, 15, 717–727.
- Dessaegne, B., & Landau, B. (2013). Interaction between language and vision: It's momentary, abstract, and it develops. *Cognition*, 127, 331–344.
- Fenson, L., Marchman, V. A., Thal, D. J., Dale, P. S., Reznick, J. S., & Bates, E. (2007). *MacArthur–Bates Communicative Development Inventories: User's guide and technical manual* (2nd ed.). Baltimore, MD: Paul H. Brookes.
- Ferreira, F., Apel, J., & Henderson, J. M. (2008). Taking a new look at looking at nothing. *Trends in Cognitive Sciences*, 12, 405–410.
- Ferry, A. L., Hespos, S. J., & Waxman, S. R. (2010). Categorization in 3- and 4-month-old infants: An advantage of words over tones. *Child Development*, 81, 472–479.
- Fitts, P. M. (1966). Cognitive aspects of information processing: III. Set for speed versus accuracy. *Journal of Experimental Psychology*, 71, 849–857.
- Folstein, J. R., Palmeri, T. J., & Gauthier, I. (2012). Category learning increases discriminability of relevant object dimensions in visual cortex. *Cerebral Cortex*, 23, 814–823.
- Frank, M. C., Sugarman, E., Horowitz, A. C., Lewis, M. L., & Yurovsky, D. (2016). Using tablets to collect data from young children. *Journal of Cognition and Development*, 17, 1–17.
- Gleitman, L., & Papafragou, A. (2012). New perspectives on language and thought. In K. J. Holyoak & R. G. Morrison (Eds.), *Oxford handbook of thinking and reasoning* (pp. 543–568). New York: Oxford University Press.
- Højsgaard, S., Halekoh, U., & Yan, J. (2006). The R package geepack for generalized estimating equations. *Journal of Statistical Software*, 15(2), 1–11.
- Hollingworth, A., Williams, C. C., & Henderson, J. M. (2001). To see and remember: Visually specific information is retained in memory from previously attended objects in natural scenes. *Psychonomic Bulletin & Review*, 8, 761–768.
- Hout, M. C., & Goldinger, S. D. (2015). Target templates: The precision of mental representations affects attentional guidance and decision-making in visual search. *Attention, Perception, & Psychophysics*, 77, 128–149.
- Johnson, P. C., Barry, S. J., Ferguson, H. M., & Müller, P. (2015). Power analysis for generalized linear mixed models in ecology and evolution. *Methods in Ecology and Evolution*, 6, 133–142.
- Jonides, J., & Gleitman, H. (1972). A conceptual category effect in visual search: O as letter or as digit. *Attention, Perception, & Psychophysics*, 12, 457–460.
- Kuwabara, M., & Smith, L. B. (2016). Cultural differences in visual object recognition in 3-year-old children. *Journal of Experimental Child Psychology*, 147, 22–38.
- Li, P., & Gleitman, L. (2002). Turning the tables: Language and spatial reasoning. *Cognition*, 83, 265–294.
- Liang, K. Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73, 13–22.
- Loewenstein, J., & Gentner, D. (2005). Relational language and the development of relational mapping. *Cognitive Psychology*, 50, 315–353.
- Lupyan, G. (2008). From chair to “chair”: A representational shift account of object labeling effects on memory. *Journal of Experimental Psychology: General*, 137, 348–369.
- Lupyan, G., & Spivey, M. J. (2010). Redundant spoken labels facilitate perception of multiple items. *Attention, Perception, & Psychophysics*, 72, 2236–2253.

- Lupyan, G., & Thompson-Schill, S. L. (2012). The evocative power of words: Activation of concepts by verbal and nonverbal means. *Journal of Experimental Psychology: General*, *141*, 170–186.
- Marchman, V. A., & Fernald, A. (2008). Speed of word recognition and vocabulary knowledge in infancy predict cognitive and language outcomes in later childhood. *Developmental Science*, *11*, F9–F16.
- Mash, C. (2006). Multidimensional shape similarity in the development of visual object classification. *Journal of Experimental Child Psychology*, *95*, 128–152.
- Morgan, P. L., Farkas, G., Hillemeier, M. M., Hammer, C. S., & Maczuga, S. (2015). 24-month-old children with larger oral vocabularies display greater academic and behavioral functioning at kindergarten entry. *Child Development*, *86*, 1351–1370.
- Oakes, L. M. (2017). Plasticity may change inputs as well as processes, structures, and responses. *Cognitive Development*, *42*, 4–14.
- Ostarek, M., & Huettig, F. (2017). Spoken words can make the invisible visible—Testing the involvement of low-level visual representations in spoken word processing. *Journal of Experimental Psychology: Human Perception and Performance*, *43*, 499–508.
- Pachella, R. G., Fisher, D. F., & Karsh, R. (1968). Absolute judgments in speeded tasks: Quantification of the trade-off between speed and accuracy. *Psychonomic Science*, *12*, 225–226.
- Pachella, R. G. (1974). The interpretation of reaction time in information-processing research. In B. H. Kantowitz (Ed.), *Human information processing: Tutorials in performance and cognition* (pp. 41–82). Oxford, UK: Lawrence Erlbaum.
- Pederson, E., Danziger, E., Wilkins, D., Levinson, S., Kita, S., & Senft, G. (1998). Semantic typology and spatial conceptualization. *Language*, *74*, 557–589.
- Perry, L. K., Samuelson, L. K., Malloy, L. M., & Schiffer, R. N. (2010). Learn locally, think globally: Exemplar variability supports higher-order generalization and word learning. *Psychological Science*, *21*, 1894–1902.
- Petersen, I. T., Bates, J. E., D'Onofrio, B. M., Coyne, C. A., Lansford, J. E., Dodge, K. A., ... Van Hulle, C. A. (2013). Language ability predicts the development of behavior problems in children. *Journal of Abnormal Psychology*, *122*, 542–557.
- R Core Team (2014). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. <<http://www.R-project.org>>.
- Rahman, R. A., & Sommer, W. (2008). Seeing what we know and understand: How knowledge shapes perception. *Psychonomic Bulletin & Review*, *15*, 1055–1063.
- Ratcliff, R. (1985). Theoretical interpretations of the speed and accuracy of positive and negative responses. *Psychological Review*, *92*, 212–225.
- Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin*, *114*, 510–532.
- Samuelson, L. K. (2002). Statistical regularities in vocabulary guide language acquisition in connectionist models and 15–20-month-olds. *Developmental Psychology*, *38*, 1016–1037.
- Smith, L. B. (2003). Learning to recognize objects. *Psychological Science*, *14*, 244–250.
- Vales, C., & Smith, L. B. (2015). Words, shape, visual search and visual working memory in 3-year-old children. *Developmental Science*, *18*, 65–79.
- Vickery, T. J., King, L. W., & Jiang, Y. (2005). Setting up the target template in visual search. *Journal of Vision*, *5*(1), 81–92.
- Vygotsky, L. S. (1980). *Mind in society: The development of higher psychological processes* (p. 32). Cambridge, MA: Harvard University Press.
- Whelan, R. (2008). Effective analysis of reaction time data. *The Psychological Record*, *58*, 475–482.
- Yoshida, H., & Smith, L. B. (2003). Known and novel noun extensions: Attention at two levels of abstraction. *Child Development*, *74*, 564–577.
- Yoshida, H., & Smith, L. B. (2005). Linguistic cues enhance the learning of perceptual cues. *Psychological Science*, *16*, 90–95.
- Zeger, S. L., & Liang, K. Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, *42*, 121–130.