Self-directed speech affects visual search performance

Gary Lupyan [a] & Daniel Swingley [b]

[a] Department of Psychology, University of Wisconsin-Madison, Madison, WI, USA

[b] Department of Psychology, Philadelphia, PA, USA

First

PLEASE SCROLL DOWN FOR ARTICLE

Psychology Press
Taylor & Francis Group

Self-directed speech affects visual search performance

Gary Lupyan[1] and Daniel Swingley[2]

[1]Department of Psychology, University of Wisconsin-Madison, Madison, WI, USA
[2]Department of Psychology, Philadelphia, PA, USA

People often talk to themselves, yet very little is known about the functions of this self-directed speech. We explore effects of self-directed speech on visual processing by using a visual search task. According to the label feedback hypothesis (Lupyan, 2007a), verbal labels can change ongoing perceptual processing—for example, actually hearing "chair" compared to simply thinking about a chair can temporarily make the visual system a better "chair detector". Participants searched for common objects, while being sometimes asked to speak the target's name aloud. Speaking facilitated search, particularly when there was a strong association between the name and the visual target. As the discrepancy between the name and the target increased, speaking began to impair performance. Together, these results speak to the power of words to modulate ongoing visual processing.

Keywords: Verbal labels; Self-directed speech; Visual search; Top-down effects; Language and thought.

Learning a language involves, among other things, learning to map object words onto categories of objects in the environment. In addition to learning that chairs are good for sitting, one learns that this class of objects has the name "chair". Clearly, such word–world associations are necessary for linguistic communication. But do hearing and producing verbal labels affect processes generally viewed as nonverbal? For example, it has been commonly observed that children spend a considerable time talking to themselves (Berk & Garvin, 1984; Vygotsky, 1962). One way to understand this seemingly odd behaviour is by considering that language is more than simply a tool for communication, but rather that it alters ongoing cognitive (and even perceptual) processing in nontrivial ways.

The idea that language alters so-called nonverbal cognition is controversial. Language is viewed by some researchers as a "transparent medium through which thoughts flow" (H. Gleitman, Fridlund, & Reisberg, 2004, p. 363), with words mapping onto concepts, but not affecting them (e.g., L. Gleitman & Papafragou, 2005; Gopnik, 2001). Although word learning is clearly constrained by nonverbal cognition, it has been argued that nonverbal cognition is not significantly influenced by learning or using words (e.g., Snedeker & Gleitman, 2004).

The alternative is that words do not simply map onto concepts, but actually change them, affecting nonverbal cognition and arguably even modulating ongoing perceptual processing. The idea that words

can affect the representations of objects to which they refer is not new. William James, for example, remarked on the power of labels to make distinctions more concrete (James, 1890, p. 333), and it has been argued that words stabilize abstract ideas in working memory, making them available for inspection (Clark, 1997; Clark & Karmiloff-Smith, 1993; Dennett, 1996; Goldstein, 1948; Rumelhart, Smolensky, McClelland, & Hinton, 1986; Vygotsky, 1962). This is not to say that different languages necessarily place strong constraints on their speakers' ability to entertain certain concepts. Rather, it is a claim that language richly interacts with putatively nonlinguistic processes such as visual processing. On this view, language is fundamentally re-entrant: Information passes in both directions, from perception/conception to linguistic encoding and from linguistic encoding back to affect "nonverbal" conceptual and perceptual representation.[1]

Insofar as performance on nonverbal tasks draws on language, interfering with language should interfere with performance on those tasks (Goldstein, 1948). Indeed, individuals with acquired language impairments (aphasia) are known to be impaired on a number of nonverbal tasks (e.g., Cohen, Kelter, & Woll, 1980; Davidoff & Roberson, 2004). Verbal interference (ostensibly, a form of down-regulation of language) has been shown to impair certain types of categorization in a strikingly similar way in healthy individuals (Lupyan, 2009). Interfering with language, even through mild articulatory suppression, also impairs healthy adults' ability to switch from one task to another (Baddeley, Chincotta, & Adlam, 2001; Emerson & Miyake, 2003; Miyake, Emerson, Padilla, & Ahn, 2004). Importantly, these specific decrements in performance due to verbal interference occur not only in relatively demanding switching tasks, but also in relatively simple and low-level perceptual tasks (e.g., Gilbert, Regier, Kay, & Ivry, 2006; Roberson & Davidoff, 2000; Roberson, Pak, &

Hanley, 2008; Winawer et al., 2007), suggesting that language actively modulates aspects of visual processing.

Results from verbal interference paradigms are difficult to interpret, in part, because it is unclear what exactly is being interfered with. An alternative way to study effects of language on perception and cognition is by implementing a dual task predicted to *increase* rather than decrease these effects. The intuition here is that whatever the influence of language in a given task, its involvement can be increased by making covert linguistic processes overt—that is, *up*-regulating language by, for example, having subjects overtly label an object or actually hear its label. Performance on these trials is then compared to performance on trials in which language is (potentially) covertly involved. A surprising finding is that when participants are asked to find a visual item among distractors, hearing its name immediately prior to searching—even when the label is entirely redundant—improves speed and efficiency of searching for the named object (or even searching among the named objects). For example, when participants searched for the numeral 2 among 5s (for hundreds of trials), actually hearing the word "two" or, in a separate experiment, hearing "ignore fives" immediately prior to searching improved overall search response times (RTs) and increased search efficiency (i.e., made the search slopes shallower; Lupyan, 2007b, 2008). Hearing an object name can also improve the ability to attend simultaneously to multiple regions of space containing the named objects (Lupyan & Spivey, 2010b) and can even make an otherwise invisible object visible (Lupyan & Spivey, 2010a). Beyond overt naming, the *meaning* ascribed to stimuli also influences visual processing. For example, Lupyan and Spivey (2008) showed that simply telling subjects that ⊓ and ⊔ should be thought of as rotated 2s and 5s dramatically improved the ability to discriminate one from the other in a visual search task (see also Risko, Dixon,

---

[1] The use of terms such as "verbal" and "nonverbal" presupposes that they are separable. On the present view, language activates (i.e., modulates) conceptual/perceptual representations with both serving as parts of an inherently interactive perceptuocognitive apparatus. A "nonverbal" representation in the present context means one that is not typically conceived as being involved in the production and comprehension of language.

Besner, & Ferber, 2006; Smilek, Dixon, & Merikle, 2006).

The present work was motivated in part by an observation from daily life: While searching specific objects, people often repeat the name of the object. Is this behaviour useful? If so, does the name simply serve as a task reminder or can it actually affect ongoing visual processing? Here, we investigated whether noncommunicative (self-directed) speech affects visual processing in the context of a search task. Participants were asked to find one or more objects among distractors. The experimental manipulation was simple: On the *speech* trials, participants were asked to actually speak the name of the target either before search (Experiment 1) or during search (Experiments 2–3). On *no-speech* trials, participants were instructed to read the name of the target object without speaking it out loud. We predicted that speaking the object's name would facilitate visual search (even though speaking during search could be seen as a distracting secondary task). We specifically sought to dissociate effects of speaking on visual processing from effects of speaking on general processes such as global attention, motivation, or a general effect of speaking on staying-on-task. If self-directed speech serves a general function of keeping participants on task (e.g., Berk & Potts, 1991), it should have the greatest facilitatory effect on trials that are most challenging—for instance, when searching for the least familiar targets, with the benefit dissipating as participants become more practised with the task. If, on the other hand, speaking helps to keep active visual representations that guide attentional processes, the effect of speaking should be largest when searching for targets having visual features most strongly associated with the label. Conversely, speaking might be detrimental when searching for objects having weaker associations with the label—for example, objects less typical of their categories or objects whose visual properties are less predictable from the label.

A useful model for thinking about the relationship between language and visual processing is one in which different levels of representation are continuously interacting (Rumelhart & McClelland, 1982; Spivey, 2008). Recognizing an object involves

not only representing its perceptual features (cf. Riesenhuber & Poggio, 2000), but combining bottom-up perceptual information with higher level conceptual information (Bar et al., 2006; Enns & Lleras, 2008; Lamme & Roelfsema, 2000). As one learns a verbal category label such as "butterfly", the label becomes associated with features that are most diagnostic or typical of the named category. With such associations in place, activation of the label—which can occur during language comprehension *or* language production—provides top-down activation of visual properties associated with the label, enhancing recognition (Lupyan & Thompson-Schill, in press).

The interaction between language and vision has, of course, been studied intensely. Hearing words has been shown to guide attention rapidly and automatically (e.g., Allopenna, Magnuson, & Tanenhaus, 1998; Andersson, Ferreira, & Henderson, 2011; Dahan & Tanenhaus, 2005; Huettig & Altmann, 2010; Salverda & Altmann, 2011; see also Anderson, Chiu, Huette, & Spivey, 2011, for review). Andersson et al. (2011), for example, showed that when viewing complex scenes, listening to naturalistic speech produces characteristic eye movement shifts (see also Richardson & Dale, 2005). In a recent analysis of distributions of saccadic launch times, Altmann (2011) demonstrated the surprising speed with which a presented word can guide overt attentional shifts: Eye movements begin to be guided toward a target as quickly as 100 ms after word onset. What has never been examined, however, is whether overtly producing speech can affect visual processing. If verbal labels modulate visual processing, then actually speaking a word out loud compared to just reading it silently may affect performance on a visual task.

EXPERIMENT 1

Participants performed a visual search, searching for a target picture among distractor pictures. Prior to each search trial, participants saw a text prompt informing them of the object they should search for. The colour of the prompt served as a cue for whether the target should be overtly verbalized.

Method

Participants

Twenty-six University of Wisconsin–Madison undergraduates (13 women) participated for course credit. Two were excluded for failing to speak out loud on speaking trials.

Materials

The targets and distractors were drawn from a set of 260 coloured drawings of familiar objects (Rossion & Pourtois, 2004). The targets were 20 pictures with the greatest values on what we call imagery concordance. This measure was computed by Rossion and Pourtois (who called it imagery) by presenting participants with a picture name (e.g., butterfly), asking them to form a mental image of the object. Then, on seeing the actual picture, participants provided a rating of concordance between their mental image and the actual picture. We chose the pictures with highest imagery-concordance values because we assumed that it would be these targets that would benefit most from being named, owing to the strong association between the label and pictorial properties (this assumption was tested explicitly in Experiment 2). The targets were: banana, barrel, baseball-bat, clothespin, envelope, fork, heart, lemon, light-bulb, nail, orange, peanut, pear, pineapple, rolling-pin, strawberry, thimble, trumpet, violin, zebra.

Procedure

Each trial began with a printed target label. On a random half of the trials, the target label was green—a cue to read it out loud. On the remaining trials, the label was presented in red, cueing participants to keep silent. After 2.2 s, the target label was replaced by the search array. Participants had to find the target by clicking on it with a computer mouse. On half of the trials, the array contained a target and 17 distractors arranged randomly on a 6×6 invisible grid. On the remaining trials, there were 35 distractors, completely filling the 6×6 array (Figure 1). Each trial had exactly one target image with the distractors drawn randomly from the 259 remaining pictures. Participants were instructed to search for a picture denoted by

the target and click on it once the picture was found. Clicking on any object ended the trial, and the response was scored correctly if the clicked object was the target.

All trial types were intermixed. Participants completed 320 trials: 20 (targets) $\times$ 2 (speech condition; speaking vs. not speaking) $\times$ 2 (distractor levels; 17 vs. 35) $\times$ 4 (blocks). A block included all Target $\times$ Speech Condition $\times$ Distractor Number combinations.

Results and discussion

Examination of audio recordings from the search trials indicated high compliance. As instructed, participants read the target name out loud on *label* trials and tended to remain silent on *no-speech* trials. Search performance was analysed using a repeated measures analysis of variance (ANOVA) with labelling condition and number of distractors as within-subject factors. RT analyses were performed on correct responses only. To avoid skewing statistical analysis with overly long response times, responses over 6 s were excluded (1.3% of trials, about 3.7 SDs above the grand mean).

Participants' performance was near ceiling ($M = 99\%$), but was nevertheless reliably higher on speaking than on no-speaking trials, $F_1(1, 23) = 5.52$, $p = .028$, Cohen's $d = 0.48$. Response speed ($M = 1,379$ ms) was likewise faster by about 50 ms when participants said the target's name out loud (Figure 2) $F_1(1, 23) = 13.27$, $p = .001$, Cohen's $d = 0.73$. Both of these differences remained significant in an item-based analysis: accuracy, $F_2(1, 19) = 6.72$, $p = .018$; RT, $F_2(1, 19) = 13.49$, $p = .002$.

Display size was a marginal predictor of errors, defined as selecting the wrong object. The error rate was somewhat higher on trials with the larger, 35-distractor, display size ($M_{errors} = 1.18\%$) than on the smaller, 17-distractor, display size ($M_{errors} = 0.63\%$), $F_1(1, 23) = 3.83$, $p = .063$, $F_2(1, 19) = 4.63$, $p = .045$. Unsurprisingly, RTs were considerably longer for the larger display size, $F_1(1, 23) = 196.01$, $p < .0005$. The speech condition by display size interaction was not reliable for either accuracy or RTs, Fs < 1.

Figure 1. *A sample search trial from Experiments 1 and 2. To view a colour version of this figure, please see the online issue of the Journal.*
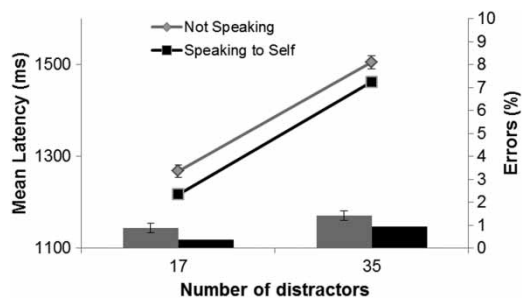

Figure 2. *Search times (line) and accuracy (bars) for Experiment 1. Error bars show ±1 standard error of the within-subject difference between the means.*

A subsequent analysis included block (1–4) as a covariate to determine whether effects of speech on visual search varied systematically during the course of the experiment. As might be expected, participants became reliably faster over the course of the experiment, $F_1(1, 23) = 5.06$, $p = .003$.

Importantly, we found a reliable interaction in accuracy between speech condition and block, $F_1(1, 23) = 7.37$, $p = .008$; $F_2(1, 19) = 6.30$, $p = .014$: The speaking advantage—despite being small in magnitude—became reliably greater over the course of the experiment. There was no reliable interaction between speech condition and block for RTs, $F < 1$.

Speaking the name of the target immediately prior to the search display made search significantly faster and more accurate. The lack of interaction between speech condition and display size indicates that search *efficiency* was not altered by speaking the name of the target (see General Discussion). That is, the benefit of speaking the name of the target may have arisen through an increase in selection confidence once the target was located, rather than any aspect of visual processing. To understand better the ways that self-directed speech influences visual search, we

conducted another experiment in which we varied aspects of the association strength between the target label and its pictorial form.

EXPERIMENT 2

Experiment 2 deviated from Experiment 1 in three ways. First, the number of elements was held constant: 1 target and 35 distractors. Second, on the speaking trials, participants were instructed to continuously speak the name of the target *during* search. This was intended to more closely approximate people's ordinary behaviour in day-to-day search situations. Third, we chose target pictures that varied in familiarity and imagery concordance in order to examine how these factors contribute to the effect of self-directed speed on visual search. We reasoned that speaking should help participants most in finding targets with strong associations between the label and the category exemplar being used for the target. Conversely, speaking may actually *hurt* performance when the target is less typical of the category.

Method

Participants
Twelve University of Pennsylvania undergraduates (7 women) participated for course credit.

Materials
The targets and distractors were drawn from the same set of images as that used in Experiment 1. For the targets, we selected 20 images having 100% picture–name agreement, but varying in familiarity and imagery concordance, as assessed by Rossion and Pourtois (2004). The target images were: airplane, banana, barn, butterfly, cake, carrot, chicken, elephant, giraffe, ladder, lamp, leaf, truck, motorcycle, mouse, mushroom, rabbit, tie, umbrella, windmill. On a given trial, any of the remaining 259 nontarget images could serve as distractors. For the item analysis, we examined the following covariates (Rossion & Pourtois, 2004): RT to name the picture, familiarity, subjective visual complexity, and imagery concordance. Familiarity was

significantly correlated with naming times, $r(18) = -.45$, $p = .04$, and visual complexity, $r(18) = -.60$, $p = .005$. No other correlations were reliable. Lexical measures included word frequency (log-transformed) from the American National Corpus (http://americannationalcorpus.org/), word length in phonemes and syllables, concreteness, and imageability obtained from the Medical Research Council (MRC) Psycholinguistic Database http://www.psy.uwa.edu.au/mrcdatabase/uwa_mrc.htm).

Procedure
Each trial began with a prompt that informed participants (a) of the object they needed to find and (b) whether they should repeat the object's name as they searched for it. For example, immediately prior to a no-speaking trial, a prompt might read: "Please search for a butterfly. Do not say anything as you search for the target." For a speaking trial, the second sentence was replaced by "Keep repeating this word continuously into the microphone until you find the target". All trial types were intermixed. Participants completed 320 trials: 20 (targets) × 2 (speech condition; speaking vs. not speaking) × 8 (blocks). A block included all Target × Speech Condition combinations.

Results and discussion

Participants showed excellent compliance with the instruction to speak the name of the target on the speaking trials and to remain silent on the no-speaking trials. The main dependent measures were accuracy and RTs to find the target. Data were analysed using a repeated measures ANOVA with speech condition as a within-subject effect and block as a covariate. All reported t tests were two-tailed.

As in Experiment 1, accuracy was extremely high ($M = 98.8\%$), revealing that subjects had no trouble remembering which item they were supposed to find and that the verbal labels were sufficiently informative to locate the correct object. Saying the object's name during search resulted in significantly higher accuracy, $M = 99.2\%$, than not repeating the name, $M = 98.4\%$, $F_1(1, 11) = 12.19$, $p = .005$,
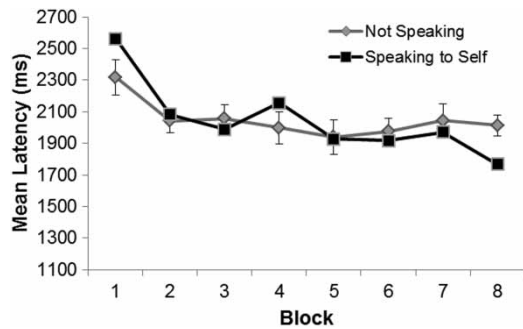
Figure 3. Response times in Experiment 2: Error bars show ±1 standard error of the within-subject difference between the means. Accuracy was significantly higher for the speaking condition throughout the task; see text.

Cohen's $d = 1.00$,[2] $F_2(1, 19) = 6.85$, $p = .017$. Participants' accuracy increased over the course of the experiment, $F(1, 11) = 10.90$, $p = .001$, but there was no reliable Speech Condition × Block interaction, $F(1, 11) = 1.49$, $p > .2$.

The analysis of RTs omitted errors and responses over 6s (3.9%). Unlike Experiment 1, there was no main effect of the speech condition on mean RTs, $F < 1$, but there was a highly reliable Speech Condition × Block interaction, $F_1(1, 11) = 8.51$, $p = .004$, $F_2(1, 19) = 9.14$, $p = .003$. As shown in Figure 3, performance on the speech trials tended to be slower than performance on no-speech trials for the initial blocks, but this pattern reversed for the latter part of the experiment.[3] For the last three blocks, participants were faster on speech trials than on no-speech trials, $F_1(1, 11) = 8.47$, $p = .014$; $F_2(1, 19) = 5.75$, $p = .027$.

We next turn to the item analysis. None of the lexical variables predicted overall search performance, but a number of characteristics of the target pictures did. The patterns of correlations are summarized in Table 1. Search was faster, $r(18) = .55$,

$p = .01$, and more accurate, $r(18) = -.54$, $p = .02$, for pictures that were visually simpler according to Rossion and Pourtois's (2004) norms. Search was faster, $r(18) = -.55$, $p = .01$, for pictures with higher imagery concordance. There was no relationship between overall accuracy and imagery concordance, $r(18) = .34$, $p = .15$. Familiarity did not predict search times or accuracy. It is apparent, glancing at Figures 2 and 3, that RTs in the present study were, controlling for display size, substantially longer than those in Experiment 1, $F_2(1, 38) = 29.09$, $p < .0005$. This difference is most likely due to the items in the present study having, by design, lower imagery-concordance values than items in Experiment 1, $F_2(1, 38) = 47.34$, $p < .0005$. Controlling for imagery concordance (a somewhat futile effort given that the values in Experiments 1 and 2 were almost nonoverlapping) showed that RTs in the present experiment were only marginally slower than those in Experiment 2, $F(1, 37) = 3.64$, $p = .064$. This analysis further demonstrates the large role that imagery concordance plays in visual search tasks of this type—a surprising finding given that participants searched for the identical target multiple times.

We next assessed which items were most affected by self-directed speech. Speaking improved accuracy most for the more familiar items, $r(18) = .51$, $p = .02$ (Figure 4, top panel; Table 1). This correlation was obtained because familiarity did not predict performance on no-speech trials, $p > .3$, but was highly predictive of performance on speaking trials, $r(18) = .55$, $p = .01$.

Finally, RTs improved marginally more for the items with the highest imagery concordance, $r(18) = .39$, $p = .08$ (Figure 4, bottom panel; Table 1). For interpretive ease, we performed a median split on the familiarity and imagery-concordance values. The label advantage ($RT_{without-speaking} - RT_{speaking}$) was reliably larger

---

[2] We repeated this and subsequent analyses of accuracy using logistic regression. In no case did these analyses provide diverging results.

[3] The speech condition by block interaction became even more reliable when we analysed the data using a linear mixed effects model that incorporated both by-subject and by-item factors (Baayen, Davidson, & Bates, 2008), $t = -3.50$, $\chi^2(1) = 12.25$, $p < .0005$. As apparent in Figure 3, the speaking advantage was greatest on Block 8 and did not reach significance for Block 6 or Block 7. However, the speech condition by block interaction remained significant even when Block 8 was excluded from the analysis, $t = -2.15$, $\chi^2(1) = 4.63$, $p = .03$. With Block 8 removed and only a single random effect, the speech condition by block interaction was marginal, $F_1(1, 11) = 3.60$, $p = .06$, $F_2(1, 11) = 3.30$, $p = .08$.

Table 1. *Summary of correlation coefficients, predicting overall performance and the self-directed speech advantage from target characteristics*

	Experiment 2			Experiment 3			
	Visual complexity	*Familiarity*	*Imagery concordance*	*Familiarity*	*Imageability*	*Typicality*	*Intracategory similarity*
Overall RT	.55*	—	−.55*	−.54*	−.67*	−.46*	−.34**
Overall accuracy (hits)	−.54*	—	—	.67*	.53*	.54*	—
Self-directed speech advantage (RT)	—	—	.39**	.51*	.44*	—	—
Self-directed speech advantage (hits)	—	.51*	—	—	—	—	.38*

Note: RT = response time. In no case did the direction of the correlation observed in RTs and accuracy contradict each other. See main text for more details.

*.0005 < p < .05. **p < .10. — = ns.

for items having imagery-concordance scores above the median, $F(1, 18) = 6.32$, $p = .022$; search items at or below the median were actually slowed by speaking, $t(10) = 2.24$, $p = .049$.[4] The label advantage in accuracy trended in the same direction, being (marginally) larger for items with above-median familiarity ratings, $F(1, 18) = 4.19$, $p = .056$.

Together these analyses suggest that speaking during search is facilitatory, but only when searching for items that are particularly familiar or have a high imagery concordance (a high level of agreement between the visual image generated based on the categorical label and the visual features of the actual target exemplar). However, the reliable speaking advantage by block interaction (Figure 3) suggests that after observing the target exemplar several times (e.g., searching for the same umbrella for the fifth time), speaking the item's name facilitated search performance. Insofar as repeated exposures strengthened the association between the label and the category exemplar, repeating the label may activate the visual properties of the target more reliably, leading to better search performance.

To summarize: Speaking facilitated search for pictures judged by independent norms to be most familiar and targets having the highest concordance between the actual image and the mental image formed by reading the name. Note that it is not

the case that any variable that facilitates search leads to facilitatory effects of self-directed speech. For example, recall that more visually complex objects took longer to find and were more likely to elicit errors. Visual complexity, however, did not predict effects of self-directed speech, $p > .5$, which is predicted, we theorize, not by general factors like search difficulty, but with the overlap between the perceptual representation activated by the label and that activated by the target item.

More than being a simple reminder, talking to oneself affected visual search performance, with the precise effect modulated by target characteristics (a fuller discussion of the labels as reminder account is presented in the General Discussion). The effect of speaking was not always facilitatory. Just as hearing a label can hurt performance when the visual quality of the item is reduced or the item is ambiguous (Lupyan, 2007a), speaking can be detrimental when the visual representation activated by the verbal label deviates from that of the target item.

EXPERIMENT 3

In Experiment 3, we attempted to generalize the effects of self-directed speech on search to a more complex "virtual shopping" task in which

---

[4] Because several items had imagery-concordance values equal to the median, the median split yielded 9 items above the median values and 11 at or below the median.
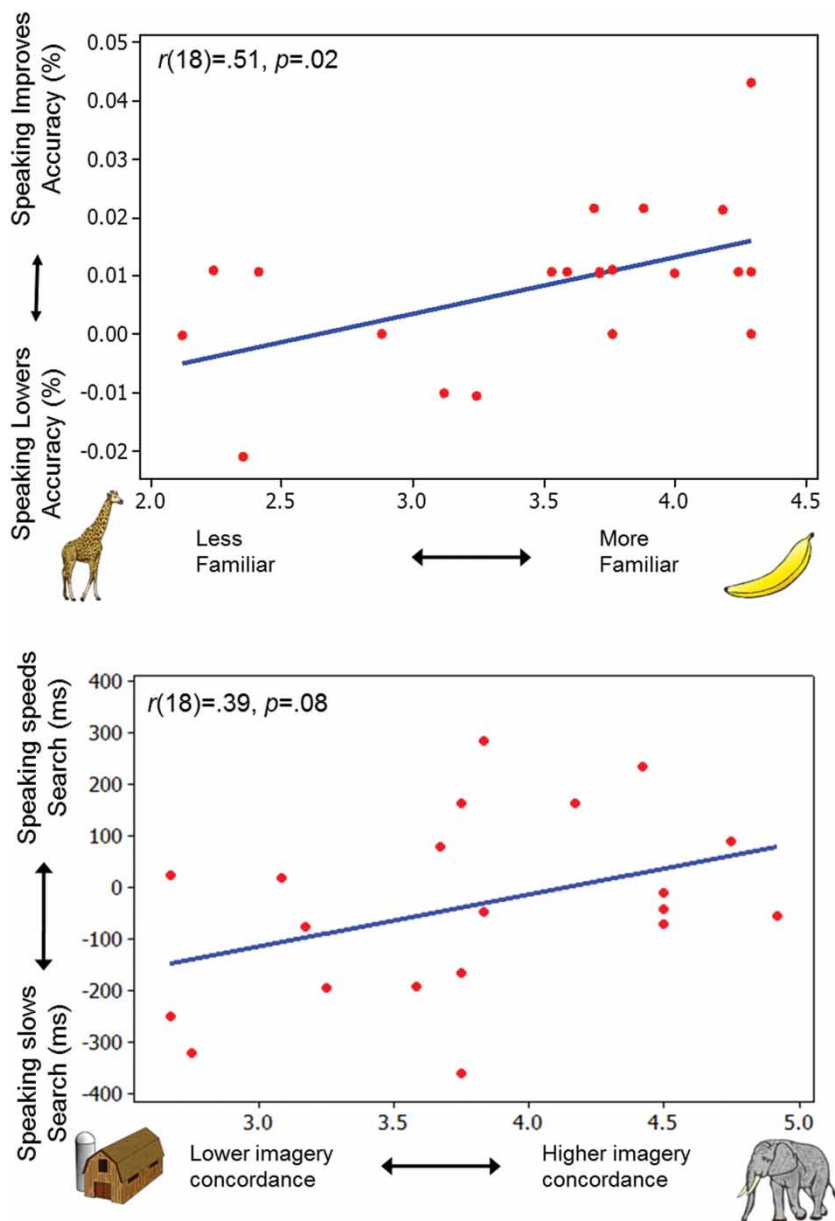
Figure 4. *Top: Relationship between item familiarity and effects of speaking on accuracy for Experiment 2. Bottom: Relationship between item imagery concordance and effects of speaking on latency for Experiment 2. The pictures show examples of items with the lowest/greatest measures for the respective predictor variables. To view a colour version of this figure, please see the online issue of the Journal.*

participants searched for supermarket products in a complex display and were required to find several rather than a single instance of each category. Including several targets per category allowed us

to examine the effect of within-category similarity on self-directed speech. The item effect observed in Experiment 1 suggested that saying a label may activate a more prototypical representation of the

item. We predicted that effects of self-directed speech would interact with within-category visual similarity such that search for visually heterogeneous targets might actually be impaired by self-produced speech insofar as it results in search more guided by category prototype.

Method

Participants
Twenty-two University of Pennsylvania undergraduates (14 women) participated for course credit.

Materials
We photographed products on supermarket shelves in the Philadelphia area and selected 30 products to serve as targets—for example, apples, Pop-Tarts, raisin bran, Tylenol, Jell-O. For each product, we obtained three pictures depicting instances of the product in various sizes and orientations. Some pictures depicted multiple instances of the product—for example, a shelf containing multiple cartons of orange juice.

Procedure
Participants were instructed to search for items while sometimes speaking the items' names. As in Experiment 2, participants were asked to repeat the name of the target category continuously during search. Each trial included all three instances of the product and 13 distractors. Clicking on an object made it disappear, thus marking it as selected. Once satisfied with their choices, participants clicked on a large "Done" button that signalled the end of the trial. To make the task more challenging, some of the distractors were categorically related to the target—for example, when searching for "Diet Coke", some distractors were of other sodas—for example, "Ginger Ale". Each subject completed 240 trials (30 targets × 8 blocks). Within each block, half the items were presented in a speech trial and half in a no-speech trial with speech and no-speech trials alternating. Across the 8 blocks, each item was presented an equal number of times in speech and no-speech conditions.

Prior to the search task, participants rated each item on typicality ("How typical is this box of Cheerios relative to boxes of Cheerios in general?") and visual quality ("How well does this picture depict a box of Cheerios?"). Participants also rated each category (e.g., the three images of Cheerios) on familiarity ("Overall, how familiar to you are the objects depicted in these pictures?") and visual similarity ("Considering only the visual appearance of these picture, how different are they from each other?"). In addition to providing us with item information, this task served to preexpose participants to all the targets. Finally, we obtained an imageability measure from a separate group of participants ($N = 28$) who were shown the written product names—for example, "Cheerios"—and were asked to rate how well they could visualize its appearance on a supermarket shelf.

Results and discussion

Participants were very accurate overall, averaging 1.5% false alarms and 97.7% hits (2.93 out of 3 targets). Trials with any misses and RTs over 10 s were excluded from the RT analysis (4.7%). Overall performance (RTs, hits, and false alarms) correlated with all four item variables (visual similarity, visual quality, familiarity, and typicality). Correlation coefficients ranged from .35 to .65 (ps between .035 and <.0005). Items that were familiar, typical, or of higher quality, and categories with greatest interitem (within-category) similarity were found faster and with higher accuracy. Of course, item characteristics were not all independent predictors—for example, familiar items and those of higher quality tended to be rated as being more typical. Typicality and familiarity measures clustered together and were not independently predictive of performance (familiarity was the stronger predictor). Within-category visual similarity predicted performance independently of familiarity; multiple regression: $F(2, 27) = 9.15$, $p = .001$.

There were no differences in RTs between the speech and no-speech conditions $M_{speech} = 2,925$ ms, $M_{no-speech} = 2,891$ ms, $F < 1$. The Speech Condition × Block interaction for RTs was qualitatively similar to that of Experiment 2, but did

not reach significance, $F_1(1, 21) = 2.43$, $p = .12$, $F_2(1, 28) = 2.12$, $p = .15$ (Figure 5). There was a small, but reliable, difference in hit rates between the two speech conditions: $M_{speech} = 97.9\%$, $M_{no\text{-}speech} = 99.1\%$, $F_1(1, 21) = 11.19$, $p = .003$, $F_2(1, 29) = 8.49$, $p = .007$, and no difference in false alarms, $M_{speech} = 1.6\%$, $M_{no\text{-}speech} = 1.5\%$, $F < 1$. While speaking, participants were more likely to miss one or more of the targets. As reported below, this apparent cost of speaking during search was modulated in interesting ways by characteristics of the target items. There was no evidence of a speed–accuracy trade-off: Search for categories yielding the longest RTs also had the most misses, $r(28) = -.67$, $p < .0005$. The speed–accuracy correlation for participants was in the same direction, but not reliable.

The item analyses in Experiment 2 suggested that effects of self-directed speech were modulated by the relationship between the item and its name. The effect of self-directed speech on RTs ($RT_{no\text{-}speech} - RT_{speech}$) in the present experiment likewise correlated with target characteristics. The effect of speaking on search RTs was mediated by familiarity, $r(28) = .51$, $p = .004$ (Table 1). As shown in Figure 6, labels tended to hurt performance for the less familiar items and improve performance for the more familiar items. Recall that in Experiment 2, speaking also improved accuracy for the most familiar items. The difference in accuracy between speaking and no-speaking trials also correlated with within-category perceptual similarity, $r(28) = -.38$, $p = .04$ (Table 1). As shown
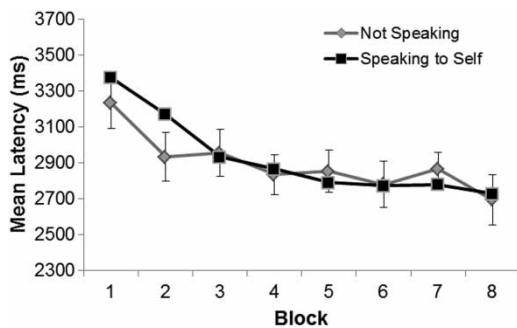
in Figure 7, speaking names of categories containing the most dissimilar items actually impaired performance. For example, for categories having below median within-category similarity scores, speaking reliably decreased accuracy, $t(14) = 3.20$, $p = .006$. Finally, the label advantage in RTs correlated positively with imageability ratings of the target category provided by a separate group of participants, $r(28) = .44$, $p = .01$ (see Table 1).

As an added demonstration that the effect of self-directed speech is modulated by target characteristics—being stronger for targets whose perceptual features are more strongly linked to their category—we divided the targets into those having characteristic colours ($N = 11$)—for example, bananas, grapes, Cheerios, raisin bran—and those items having weaker associations with a specific colour—for example, Jell-O, Pop-Tarts. The speaking advantage was greater for colour-diagnostic items—for which speaking significantly improved RTs—than for non-colour-diagnostic items—for which speaking marginally increased RTs: colour diagnosticity by speech condition interaction, $F(1, 28) = 7.35$, $p = .01$.

Finally, we observed in Experiment 3 a curious gender difference in performance. Men had a significantly lower hit rate, $F(1, 20) = 5.02$, $p = .037$, and were significantly slower to find the targets, $F(1, 20) = 6.37$, $p = .02$, than women. The gender effect on RTs was substantial: Men took on average 350 ms longer per trial. This effect was replicated in an item analysis, $F_2(1, 29) = 43.40$, $p < .0005$ (the only item on which men were faster than women was "Degree Deodorant"). There was a marginal Gender × Speech Condition interaction for hit rates, $F(1, 20) = 3.79$, $p = .066$: Self-directed speech decreased overall performance slightly more for men than for women. An examination of item ratings revealed that there were no gender differences in subjective ratings of familiarity, visual quality, or visual similarity, $Fs < 1$; the greatest gender difference was obtained in judgements of typicality (men compared to women rated the items as being less typical); however, these differences did not reach significance, $F(1, 20) = 2.66$, $p = .12$. There were no gender differences in Experiments 1 or 2, $Fs < 1$. It is unclear whether
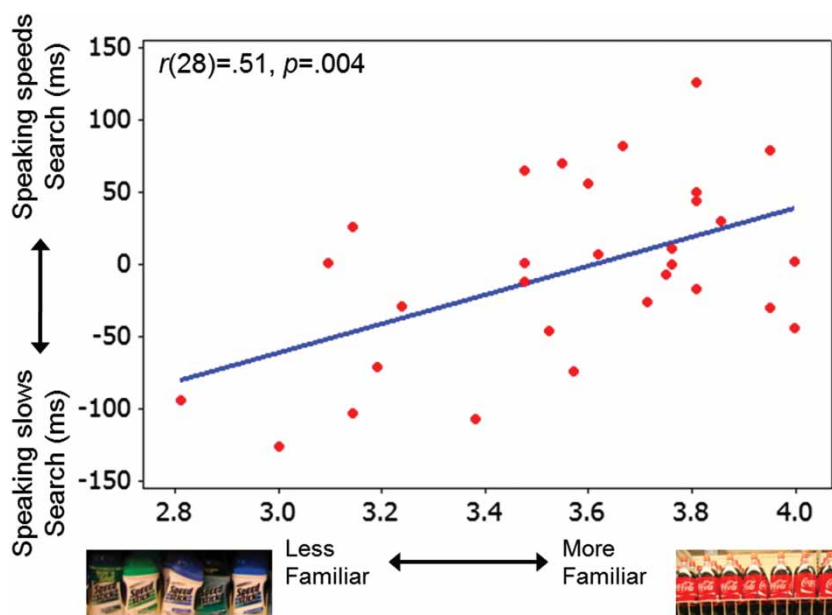

Figure 5. *Response times in Experiment 3: Error bars show ± 1 standard error of the within-subject difference between the means.*

Figure 6. *Relationship between familiarity and effects of speaking on response times for Experiment 3. The pictures show examples of items with the lowest/greatest measures for the respective predictor variables. To view a colour version of this figure, please see the online issue of the Journal.*

this gender difference (which was replicated in a study not described here) arises from our choice of materials or from having to select multiple targets per trial (recall that in Experiments 1–2 there was only a single target per trial).

Despite some differences in the main effects between the two studies, Experiment 3 supported the findings of Experiment 2 with a larger, more perceptually varied and true-to-life materials. As in Experiment 2, speaking aided search for the more familiar and imageable items (see Table 1). In contrast to Experiment 2, overall accuracy (hit rate) was actually lower on speaking trials. The reduced accuracy was greater for items having low within-category similarity. This finding is consistent with the idea that speaking an object name activates a category representation that best matches (proto) typical exemplars (Lupyan & Thompson-Schill, in press). When the task requires finding items that have less typical features, and when participants need to find visually heterogeneous items from the same category, speaking can impair performance.

GENERAL DISCUSSION

In this work, we examined effects of self-directed speech on performance on a simple visual task. Speaking the name of the object for which one was searching affected performance on the visual search task relative to intermixed trials on which participants read the word but did not actually speak it before or during search. The effect of speaking depended strongly on the characteristics of the target item. Search was improved for the most familiar and prototypical items—those for which speaking the name is hypothesized to evoke the visual representation that best matches the visual characteristics of the target item (Lupyan, 2008; Lupyan & Spivey, 2010b). Search was unaffected or impaired as the discrepancy between the name and target—measured by measures of familiarity and imagery concordance —was increased.

Facilitation due to speaking also became larger with repeated exposures to the target items.
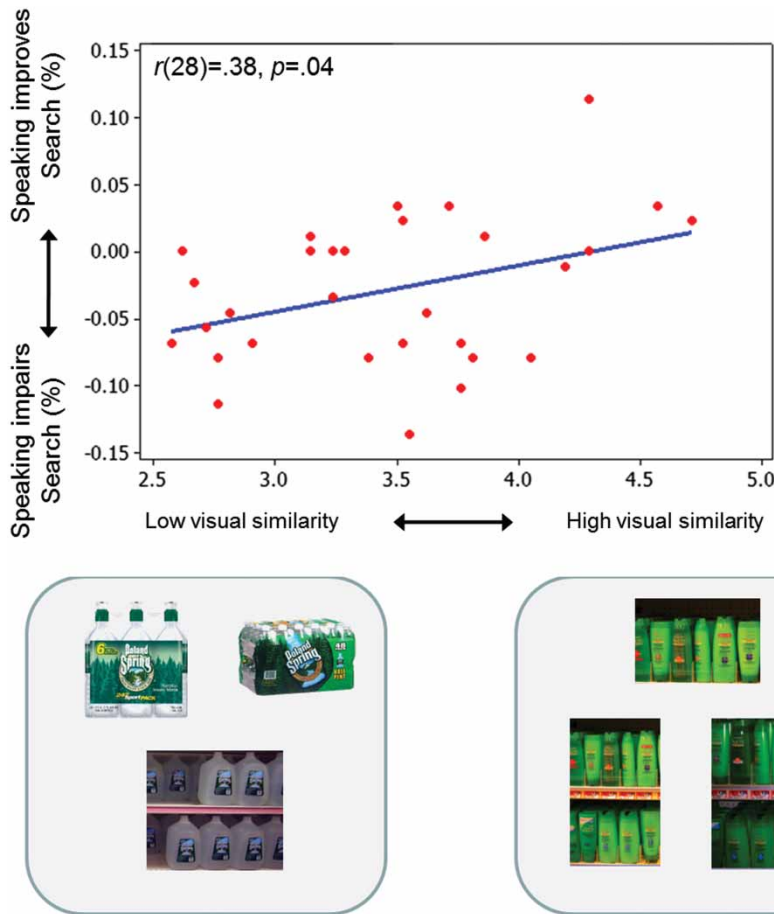
Figure 7. *The relationship between within-category visual similarity and effects of speaking on hit rate in Experiment 3. Poland Spring Water and Fructis Shampoo were, respectively, the categories with the least and the most within-category visual similarity. To view a colour version of this figure, please see the online issue of the Journal.*

Arguably this occurred because multiple exposures strengthened the associations between the label (e.g., "elephant") and the visual exemplar (a given picture of an elephant; Lupyan, Thompson-Schill, & Swingley, 2010). The idea that saying a category name activates a more prototypical representation of the category is also supported by the finding that speaking the name actually hurts performance for items with low within-category similarity. One implication is that repeating the word "knife" may, for example, help an airport baggage screener spot typical knives, but actually make it more difficult to find less prototypical knives.

On our view, the reason speaking the target name affects visual search performance is that speaking its name helps to activate and/or keep active visual (as well as nonvisual) features that are diagnostic of the object's category, facilitating the processing of objects with representations overlapping those activated by the label (Lupyan, 2008; Lupyan & Thompson-Schill, in press; see also Soto & Humphreys, 2007, for a related proposal). This activation of visual features occurs during silent reading as well. Indeed, it is what allows foreknowledge of the target to guide search (e.g., Vickery, King, & Jiang, 2005). Self-directed

speech, as implemented in the present studies, is hypothesized to further enhance this process.

An important question is whether self-directed speech affects the process of locating the target per se, or only aids in identifying it once it is located (e.g., see Castelhano, Pollatsek, & Cave, 2008, for a similar argument regarding the role of target typicality in search).[5] In the present case, it is admittedly difficult to disentangle an effect of self-directed speech on search guidance from its effect on target identification. The failure to find an interaction between speaking condition and display size in Experiment 1 suggests that speaking the name of the target does not help initially locating it, relative to just reading the target name. This is in contrast to earlier studies showing that hearing a category label prior to search can improve search efficiency (Lupyan, 2007b, see also 2008). A direct comparison is difficult because these earlier studies used much simpler visual forms and required target presence/absence responses rather than actually selecting the target. Also, in contrast to word cues, which could be presented with high temporal precision, we did not have precise control here over the timing of participants' self-produced speech. Slight differences in the timing of the word relative to the onset of the search display could be important[6]: The effects of hearing labels on visual processing have been found to have a characteristic time-course, peaking about 0.5–1.5 seconds after the presentation of the label and declining afterwards (Lupyan & Spivey, 2010b). In summary, although the present results provide evidence that self-directed speech affects some aspect of the visual search process that is specific to the target category, there is no evidence at present that self-directed speech affected the efficiency of locating the target.

An important remaining question is whether effects of speaking on visual search arise from the act of production itself or from hearing one's speech. Although this distinction is of little practical importance (one almost always hears oneself speak), a full understanding of the mechanism by which speech interacts with visual processing requires the two explanations to be teased apart (Huettig & Hartsuiker, 2010). One way to do this would be to compare speaking aloud and silent mouthing. The prediction is that silent mouthing will result in performance in between silent reading and vocalizing (see also MacLeod, Gopie, Hourihan, Neary, & Ozubko, 2010, for effects of overt speaking on recognition memory). However, regardless of whether it is the production or the subsequent perception of one's speech that affects visual search performance, the important message of the present results is that not only can externally provided linguistic cues affect visual processing, but self-produced language can function in some of the same ways.

Distilling the mechanisms by which word production affects visual processing clearly requires further work, but the observed pattern of results places some constraints on possible mechanisms. We highlight three alternatives to our position that self-directed speech activated visual properties of the target category over and above silently reading the word. We believe these alternatives are not well supported by the pattern of results.

1. *Self-directed speech affects only the cognitive process of selecting the target, not the visual process of recognizing it.* Given that self-directed speech does not affect search efficiency, there is a possibility that self-directed speech affected the selection of the target rather than any processing involved in visual recognition of the target. The best evidence against this possibility is that the effect of self-directed speech was modulated by target

---

[5] There is evidence that conceptual characteristics such as typicality of the target do affect visual guidance. For example, Zelinsky and colleagues (Alexander, Zhang, & Zelinsky, 2010; Schmidt & Zelinsky, 2009; Yang & Zelinsky, 2009) have found that following a verbal description of the target, participants are more likely to move their eyes to a category typical target. Additionally, studies using the visual world paradigm have consistently shown that hearing words activates both visual and nonvisual information, which rapidly affects eye movements (Dahan & Tanenhaus, 2005; Huettig & Altmann, 2007; Yee & Sedivy, 2006).

[6] Recordings of participants' speech from the present work revealed a wide variability in the onset, speed, and duration of self-directed speech. A post hoc analysis of voice recordings during the speech trials failed to find reliable correlations between search times and onset or offset times of self-directed speech.

characteristics (see Figures 4, 6, 7, Table 1). This suggests that self-directed speech affected the identification of the target (as distinct from, for example, affecting a global parameter such as the threshold for target selection). In addition, recent work aimed specifically at exploring the effect of hearing object names on visual processing has shown that hearing completely redundant verbal labels affects deployment of attention even when identification of the target is not required (Lupyan & Spivey, 2010b; Lupyan & Thompson-Schill, in press), although it is possible that such effects would not be observed for self-produced labels. A discussion of the relationship between linguistic effects on visual processing and theories of visual search can be found in Lupyan and Spivey (2010b).

2. *Self-directed speech helps subjects to remember what they are searching for.* On this so-called labels-as-reminder account, speaking helped participants to remember what they were looking for, or kept participants on task. Clearly such rehearsal is a useful strategy for remembering a list of items, but we do not think that effects of labels in the present studies had a significant impact on participants' memory for a single word. Although it is possible that the small (but reliable) accuracy boost on speech trials in Experiment 1 was due to reducing the (already very low) probability of forgetting what the target was, the labels-as-reminder account does not predict any of the correlation patterns between target picture characteristics and effects of speaking on search times and accuracies (see Table 1) or the interactions between block and label effect in Experiments 1 and 2 (the finding that the facilitatory effect of the label increased during the course of the task). Indeed, the labels-as-reminder account might predict the opposite: Performance in the task became easier as participants became more practised (as evidenced by shorter RTs), and hence presumably participants should benefit less from any memory aids. In contrast, the observed correlations are expected on an account in which an increase in association between a label and a visual image increases the effectiveness of the label in activating visual properties of the image (Lupyan, 2007a; Lupyan et al., 2010). Finally, the labels-as-reminder explanation also does not predict why speaking lowered performance in Experiment 3 for certain categories, particularly those with items having low visual similarity. On our account, this is obtained because saying a category name may activate a more prototypical representation of the category, making it more difficult to locate all the members of a visually heterogeneous category.

3. *Self-directed speech helps via word-to-word matching.* On this account, self-directed speech affected visual search by facilitating the mapping between the name of the target and the name of objects in the search array (assuming that those names are rapidly activated upon seeing the objects). This alternative interpretation of the results rests on two assumptions. The first is that pictures rapidly and automatically activate their names. This assumption has support in the literature (Zelinsky & Murphy, 2000; see also Meyer, Belke, Telling, & Humphreys, 2007; cf. Jescheniak, Schriefers, Garrett, & Friederici, 2002). The second assumption is that the target location process involves matching the name of the target to names generated by the pictures in the display. We cannot conclusively rule this out, but it strikes us as unlikely that such a name-matching procedure can be performed for 36 pictures in ~1.5 seconds (Figure 2).

The present work is the first to examine effects of self-directed speech in a relatively simple visual task, adding to the growing literature showing that language serves a number of extracommunicative functions and, under some conditions, has the power to modulate visual processes (see also Lupyan & Spivey, 2010a). In line with Vygotsky's claim (1962) that the function of self-directed speech extends beyond verbal rehearsal (see also Baddeley et al., 2001; Carlson, 1997), we view the present results as an added demonstration that language not only is a communicative tool,

but modulates ongoing cognitive and perceptual processes in the language user, thus affecting performance on nonlinguistic tasks.

REFERENCES

Alexander, R., Zhang, W., & Zelinsky, G. J. (2010). Visual similarity effects in categorical search. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 1222–1227). Austin, TX: Cognitive Science Society.

Allopenna, P., Magnuson, J., & Tanenhaus, M. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, 38(4), 419–439.

Altmann, G. T. M. (2011). Language can mediate eye movement control within 100 milliseconds, regardless of whether there is anything to move the eyes to. *Acta Psychologica*, 137(2), 190–200. doi:10.1016/j.actpsy.2010.09.009

Anderson, S. E., Chiu, E., Huette, S., & Spivey, M. J. (2011). On the temporal dynamics of language-mediated vision and vision-mediated language. *Acta Psychologica*, 137(2), 181–189. doi:10.1016/j.actpsy.2010.09.008

Andersson, R., Ferreira, F., & Henderson, J. M. (2011). I see what you're saying: The integration of complex speech and scenes during language comprehension. *Acta Psychologica*, 137(2), 208–216. doi:16/j.actpsy.2011.01.007

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390–412. doi:16/j.jml.2007.12.005

Baddeley, A. D., Chincotta, D., & Adlam, A. (2001). Working memory and the control of action: Evidence from task switching. *Journal of Experimental Psychology: General*, 130(4), 641–657. doi:10.1037/0096-3445.130.4.641

Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmidt, A. M., Dale, A. M., et al. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences of the United States of America*, 103(2), 449–454. doi:10.107/pnas.0507062103

Berk, L. E., & Garvin, R. A. (1984). Development of private speech among low-income Appalachian children. *Developmental Psychology*, 20(2), 271–286.

Berk, L. E., & Potts, M. K. (1991). Development and functional-significance of private speech among attention-deficit hyperactivity disordered and normal boys. *Journal of Abnormal Child Psychology*, 19(3), 357–377.

Carlson, R. A. (1997). *Experienced cognition* (1st ed.). Hove, UK: Psychology Press.

Castelhano, M. S., Pollatsek, A., & Cave, K. R. (2008). Typicality aids search for an unspecified target, but only in identification and not in attentional guidance. *Psychonomic Bulletin & Review*, 15(4), 795–801.

Clark, A. (1997). *Being there: Putting brain, body, and world together again*. Cambridge, MA: MIT Press.

Clark, A., & Karmiloff-Smith, A. (1993). The cognizer's innards: A psychological and philosophical perspective on the development of thought. *Mind & Language*, 8(4), 487–519.

Cohen, R., Kelter, S., & Woll, G. (1980). Analytical competence and language impairment in aphasia. *Brain and Language*, 10(2), 331–347.

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.

Davidoff, J., & Roberson, D. (2004). Preserved thematic and impaired taxonomic categorisation: A case study. *Language and Cognitive Processes*, 19(1), 137–174.

Dennett, D. C. (1994). The role of language in intelligence. In J. Khalfa (Ed.), *What is intelligence?* The Darwin College Lectures, Cambridge.

Emerson, M. J., & Miyake, A. (2003). The role of inner speech in task switching: A dual-task investigation. *Journal of Memory and Language*, 48(1), 148–168.

Enns, J. T., & Lleras, A. (2008). What's next? New evidence for prediction in human vision. *Trends in Cognitive Sciences*, 12(9), 327–333. doi:10.1016/j.tics.2008.06.001

Gilbert, A. L., Regier, T., Kay, P., & Ivry, R. B. (2006). Whorf hypothesis is supported in the right visual field but not the left. *Proceedings of the National Academy of Sciences of the United States of America*, 103(2), 489–494.

Gleitman, H., Fridlund, A. J., & Reisberg, D. (2004). *Psychology* (6th ed.). New York, NY: Norton & Company.

Gleitman, L., & Papafragou, A. (2005). Language and thought. In K. Holyoak & B. Morrison (Eds.),

Cambridge handbook of thinking and reasoning (pp. 633–661). Cambridge, UK: Cambridge University Press.

Goldstein, K. (1948). *Language and language disturbances*. New York, NY: Grune & Stratton.

Gopnik, A. (2001). Theories, language, and culture: Whorf without wincing. *Language acquisition and conceptual development* (pp. 45–69). Cambridge, UK: Cambridge University Press.

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. doi:10.1080/13506280601130875

Huettig, F., & Altmann, G. T. M. (2010). Looking at anything that is green when hearing "frog": How object surface colour and stored object colour knowledge influence language-mediated overt attention. *The Quarterly Journal of Experimental Psychology*. doi:10.1080/17470218.2010.481474

Huettig, F., & Hartsuiker, R. J. (2010). Listening to yourself is like listening to others: External, but not internal, verbal self-monitoring is based on speech perception. *Language and Cognitive Processes*, *25*(3), 347. doi:10.1080/01690960903046926

James, W. (1890). *Principles of psychology* (Vol. 1). New York, NY: Holt.

Jescheniak, J. D., Schriefers, H., Garrett, M. F., & Friederici, A. D. (2002). Exploring the activation of semantic and phonological codes during speech planning with event-related brain potentials. *Journal of Cognitive Neuroscience*, *14*(6), 951–964. doi:10.1162/089892902760191162

Lamme, V. A. F., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences*, *23*(11), 571–579.

Lupyan, G. (2007a). *The label feedback hypothesis: Linguistic influences on visual processing* (unpublished PhD thesis). Pittsburgh, PA: Carnegie Mellon University.

Lupyan, G. (2007b). Reuniting categories, language, and perception. In D. S. Mcnamara & J. G. Trafton (Eds.), *Twenty-Ninth Annual Meeting of the Cognitive Science Society* (pp. 1247–1252). Austin, TX: Cognitive Science Society.

Lupyan, G. (2008). The conceptual grouping effect: Categories matter (and named categories matter more). *Cognition*, *108*, 566–577.

Lupyan, G. (2009). Extracommunicative functions of language: Verbal interference causes selective categorization impairments. *Psychonomic Bulletin & Review*, *16*(4), 711–718. doi:10.3758/PBR.16.4.711

Lupyan, G., & Spivey, M. J. (2008). Perceptual processing is facilitated by ascribing meaning to novel stimuli. *Current Biology*, *18*(10), R410–R412.

Lupyan, G., & Spivey, M. J. (2010a). Making the invisible visible: Auditory cues facilitate visual object detection. *PLoS ONE*, *5*(7), e11452. doi:10.1371/journal.pone.0011452

Lupyan, G., & Spivey, M. J. (2010b). Redundant spoken labels facilitate perception of multiple items. *Attention, Perception & Psychophysics*, *72*(8), 2236–2253. doi:10.3758/APP.72.8.2236

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*(1), 170–186. doi:10.1037/a0024904

Lupyan, G., Thompson-Schill, S. L., & Swingley, D. (2010). Conceptual penetration of visual processing. *Psychological Science*, *21*(5), 682–691.

MacLeod, C. M., Gopie, N., Hourihan, K. L., Neary, K. R., & Ozubko, J. D. (2010). The production effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *36*(3), 671–685. doi:10.1037/a0018785

Meyer, A. S., Belke, E., Telling, A. L., & Humphreys, G. W. (2007). Early activation of object names in visual search. *Psychonomic Bulletin & Review*, *14*(4), 710–716.

Miyake, A., Emerson, M. J., Padilla, F., & Ahn, J. C. (2004). Inner speech as a retrieval aid for task goals: The effects of cue type and articulatory suppression in the random task cuing paradigm. *Acta Psychologica*, *115*(2–3), 123–142. doi:10.1016/j.actpsy.2003.12.004

MRC Psycholinguistic Database. Retrieved September 7, 2011, from http://www.psy.uwa.edu.au/mrcdatabase/uwa_mrc.htm.

Richardson, D. C., & Dale, R. (2005). Looking to understand: The coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension. *Cognitive Science*, *29*, 1045–1060.

Riesenhuber, M., & Poggio, T. (2000). Models of object recognition. *Nature Neuroscience*, *3*(Suppl), 1199–1204. doi:10.1038/81479

Risko, E. F., Dixon, M. J., Besner, D., & Ferber, S. (2006). The ties that keep us bound: Top-down influences on the persistence of shape-from-motion. *Consciousness and Cognition*, *15*(2), 475–483. doi:16/j.concog.2005.11.004

Roberson, D., & Davidoff, J. (2000). The categorical perception of colors and facial expressions: The effect of verbal interference. *Memory & Cognition*, *28*(6), 977–986.

Roberson, D., Pak, H., & Hanley, J. R. (2008). Categorical perception of colour in the left and right visual field is verbally mediated: Evidence from Korean. *Cognition*, *107*(2), 752–762. doi:10.1016/j.cognition.2007.09.001

Rossion, B., & Pourtois, G. (2004). Revisiting Snodgrass and Vanderwart's object pictorial set: The role of surface detail in basic-level object recognition. *Perception*, *33*(2), 217–236.

Rumelhart, D. E., & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: 2. The contextual enhancement effect and some tests and extensions of the model. *Psychological Review*, *89*(1), 60–94.

Rumelhart, D. E., Smolensky, D., McClelland, J. L., & Hinton, G. E. (1986). Parallel distributed processing models of schemata and sequential thought processes. *Parallel distributed processing* (Vol. II, pp. 7–57). Cambridge, MA: MIT Press.

Salverda, A. P., & Altmann, G. T. M. (2011). Attentional capture of objects referred to by spoken language. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(4), 1122–1133. doi:10.1037/a0023101

Schmidt, J., & Zelinsky, G. J. (2009). Search guidance is proportional to the categorical specificity of a target cue. *Quarterly Journal of Experimental Psychology*, *62*(10), 1904–1914. doi:10.1080/17470210902853530

Smilek, D., Dixon, M. J., & Merikle, P. M. (2006). Revisiting the category effect: The influence of meaning and search strategy on the efficiency of visual search. *Brain Research*, *1080*, 73–90.

Snedeker, J., & Gleitman, L. (2004). Why is it hard to label our concepts? In D. G. Hall & S. R. Waxman (Eds.), *Weaving a lexicon* (illustrated ed., pp. 257–294). Cambridge, MA: MIT Press.

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), 730–737. doi:10.1037/0096-1523.33.3.730

Spivey, M. J. (2008). *The continuity of mind*. New York, NY: Oxford University Press.

The American National Corpus. Retrieved September 7, 2011, from http://www.anc.org/.

Vickery, T. J., King, L.-W., & Jiang, Y. (2005). Setting up the target template in visual search. *Journal of Vision*, *5*(1), 81–92. doi:10:1167/5.1.8

Vygotsky, L. (1962). *Thought and language*. Cambridge, MA: MIT Press.

Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(19), 7780–7785.

Yang, H., & Zelinsky, G. J. (2009). Visual search is guided to categorically-defined targets. *Vision Research*, *49*(16), 2095–2103. doi:10.1016/j.visres.2009.05.017

Yee, E., & Sedivy, J. C. (2006). Eye movements to pictures reveal transient semantic activation during spoken word recognition. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *32*(1), 1–14. doi:10.1037/0278–7393.32.1.1

Zelinsky, G. J., & Murphy, G. L. (2000). Synchronizing visual and language processing: An effect of object name length on eye movements. *Psychological Science*, *11*(2), 125–131.