Eye Movements and the Label Feedback Effect:

Speaking Modulates Visual Search, but Probably Not Visual Perception

by

Katherine P. Hebert

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Arts

Approved July 2016 by the
Graduate Supervisory Committee:

Stephen D. Goldinger, Chair
Corianne Rogalsky
Samuel M. McClure

ARIZONA STATE UNIVERSITY

August 2016

ABSTRACT

The label-feedback hypothesis (Lupyan, 2007) proposes that language can modulate low- and high-level visual processing, such as "priming" a visual object. Lupyan and Swingley (2012) found that repeating target names facilitates visual search, resulting in shorter reaction times (RTs) and higher accuracy. However, a design limitation made their results challenging to assess. This study evaluated whether self-directed speech influences target locating (i.e. attentional guidance) or target identification after location (i.e. decision time), testing whether the Label Feedback Effect reflects changes in visual attention or some other mechanism (e.g. template maintenance in working memory). Across three experiments, search RTs and eye movements were analyzed from four within-subject conditions. People spoke target names, nonwords, irrelevant (absent) object names, or irrelevant (present) object names. Speaking target names weakly facilitates visual search, but speaking different names strongly inhibits search. The most parsimonious account is that language affects target maintenance during search, rather than visual perception.

ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. Stephen Goldinger, for his continued expert guidance throughout the research process, as well as my committee members for their insightful comments and suggestions. This project would also not have been possible without the tremendous programming assistance and patience of Stephen Walenchok.

TABLE OF CONTENTS

Page

LIST OF FIGURES

You're late for work, and your keys are nowhere to be found. This scenario should not be too difficult for most of us to imagine. You mutter the word "keys" aloud to yourself while you search, and ultimately find them flung on the ground next to a pair of shoes. Did repeating the word "keys" out loud to yourself help you to detect your keys faster? The Label Feedback Hypothesis (Lupyan, 2007b; Lupyan & Swingley, 2012) proposes that hearing or speaking the name of a target object aides in target detection. The idea that categories and concepts acquired through language learning may impact human perception is not new, and has intrigued scientists since the 1950's (Bruner, Goodnow, & Austin, 1956; Whorf, 1956). A bulk of the research on the subject is in the area of Categorical Perception, the phenomenon wherein people are better at discriminating between stimuli of different categories than they are at discriminating between members of the same category (Goldstone, 1994; Harnad, 1987). The past decade has seen a large influx in research on the top-down influences of language on visual and auditory perception, investigating whether language experience shapes even "low-level" perception, and the results have been mixed and therefore controversial (e.g. Boutonnet & Lupyan, 2015; Klemfuss, Prinzmetal, & Ivry, 2012; Meteyard, Bahrami, & Vigliocco, 2007; Suegami, Aminihajibashi, & Laeng, 2013).

Some researchers maintain that while words map onto concepts, they do not influence them, and nonverbal cognition is likewise not affected by learning or using words (e.g. Gleitman & Papafragou, 2005; Gopnik, 2001; Snedeker & Gleitman, 2004). The alternative is that nonverbal processes are directly impacted by language, and that words may actually change concepts and drive perception (Deutscher, 2010; Gilbert, Regier, Kay, & Ivry, 2006; Regier & Kay, 2009). A predominant way to explore this question has been to utilize verbal-interference paradigms, using language to disrupt performance on distinctly nonverbal, low-level perceptual tasks (where nonverbal interference, conversely, has no impact on performance) (e.g. Drivonikou et al., 2007; Gilbert, Regier, Kay, & Ivry, 2006; Lupyan, 2009). Some studies, however, have taken the opposite approach, aiming to use language to *increase* performance on nonverbal tasks (e.g. Risko, Dixon, Besner, & Ferber, 2006; Smilek, Dixon, & Merikle, 2006; Soto, & Humphreys, 2007). Visual search tasks (looking for a pair of keys among shoes, jackets and backpacks, for example) are often utilized here, and have shown that hearing the name of that target improves speed, accuracy, and efficiency in finding it among distractors (Lupyan, 2007a, 2008; Lupyan & Spivey, 2010a), and can also make an otherwise invisible object salient (Lupyan & Spivey, 2010b; Lupyan & Ward, 2013).

More recent research has revolved around The Label Feedback Hypothesis, examining whether self-directed speech impacts search performance. Lupyan and Swingley (2012) had participants either speak the name of the target item aloud to themselves during visual search, or else not speak at all. They found that speaking facilitated search, resulting in shorter reaction times (RTs) and higher accuracy, especially when target familiarity and imagery concordance were high (Figure 1). But subjects were at ceiling (99% mean accuracy), and their paradigm can be improved upon. They also left key questions unanswered, including whether speaking the name simply serves as a task reminder, or whether the verbal process actually modulates ongoing visual processing. Only two conditions were compared, speaking target names versus silence, which means comparing a single-task to a dual task. Additionally, there is nothing to ensure that participants do not silently rehearse the target name during the non-speaking trials. Lastly, by using only search times, they could not evaluate whether self-directed speech influences target locating (i.e. attentional guidance) or target identification after location (i.e. decision time). In this study, we aimed to determine whether the Label Feedback Effect reflects changes in visual attention, or some other mechanism (e.g. template maintenance in working memory) by using both search RTs and eye movements.

To allow more direct comparisons, we had participants simultaneously speak while performing visual search in all conditions, varying the content of their speech. Four within-subjects, blocked conditions were tested: 1) Participants repeated target names during search, as in Lupyan and Swingley (Target condition), 2) Participants repeated nonwords during search (Nonword condition), 3) Participants repeated the names of real objects, not in the display (Distractor Absent condition), and 4) Participants repeated the names of distractor objects that *were* present on-screen (Distractor Present condition). The Target condition was expected to show most efficient search, reflecting improved guidance and/or perceptual classification. The Nonword condition served as control, allowing us to equate the challenges of performing a dual-task, without true linguistic content. The Distractor Absent condition activated true object labels, but no visual distraction of those competing objects on-screen. Therefore, attention could not be guided to those objects, nor would perceptual classification be challenged by fixating repeatedly named non-target items. Finally, the Distractor Present condition entailed both challenges, as language could drive attention to the named distractor, and could make perceptual decisions more difficult. These conditions gave a wider range of not only search facilitation but also interference due to verbal processes, allowing us to better disentangle the effects Lupyan and Swingley (2012) observed. Three experiments were conducted, two variations of a behavioral study and an eye-tracking study. Experiments 1 and 2 had the exact same paradigm, but the latter was conducted on a much larger display, to lengthen eye movements and potentially amplify any differences among the four conditions.

For the eye-tracking portion of the project, there were two foreseeable outcomes that would be of interest. One possibility was that eye movements reveal more efficient search patterns in the conditions where speech facilitates search, and less efficient search when speech interferes with search. The other option was that search patterns are of a consistent efficiency across all conditions, and the impact on RTs take place in the space of time between when the target has been found and when the keyboard response is made (i.e. decision time). The Label Feedback Hypothesis predicts that language impacts the allocation of attention to the environment. If language does not influence search efficiency, but it *does* influence the ability to identify a target, it would have significant theoretical ramifications, strongly implicating an impact of language on low-level perceptual processes. Regardless of the outcome observed, the present paradigm offered a collection of manipulations that provided the opportunity to explore both verbal and attention processes, separately and together, to better illuminate the extent to which speech impacts perception in visual search.

METHODS

Experiment 1

The purpose of Experiment 1 was to determine whether there is in fact a label feedback effect, and allowed us to better disentangle the details of the effect with eye tracking (in Experiment 3).

Participants. 59 participants were recruited from the Arizona State University Psychology 101 subject pool. The participants were given course credit for their participation. All were native English speakers and had normal or corrected-to-normal vision by self-report. Five subjects were excluded from data analysis based on performance—outliers were identified as anyone whose average RTs or accuracies were two or more standard deviations above the group mean on any of the four visual search conditions.

Apparatus. Stimuli were presented using Dell computers and 16-inch monitors, and participants responded via keyboard. Data was collected on up to two computers simultaneously, each sharing identical hardware and software profiles. Each computer was in its own room, so that participants could not hear each other speaking. The experiment was administered using E-Prime 2.0 stimulus presentation software (Schneider, Eschman, & Zuccolotto, 2012).

Stimuli and design. Across all four conditions, the target object words and the distractor object words were either one, two, or three syllables (approximately 25% were one syllable, 50% were two syllables, and 25% were three syllables). The mono- and bisyllable nonwords for the Nonword condition were borrowed exactly from Goldinger (1998); trisyllable nonwords involved adding prefixes or suffixes onto a bisyllable word from Goldinger's list. The majority of object pictures came from the "Massive Memory" database (Brady, Konkle, Alvarez, & Oliva, 2008; Konkle, Brady, Alvarez, & Oliva, 2010, cvcl.mit.edu/MM/stimuli.html). A small number of pictures came from Google Image searches. All images were approximately 100x100 pixels. This set of pictures gave us 192 unique target objects, and approximately 2,000 distractors. Similarity ratings from a multidimensional scaling database were utilized for object categories, and conflicting objects/categories were never paired (Hout, Goldinger, & Brady, 2014). There were four conditions (Target Condition, Nonword Condition, Distractor Absent Condition, and Distractor Present Condition), each randomized by block, all within-subjects. In each condition, the participant repeated a word out loud to herself during search. The four blocks were presented in random order. Each block consisted of 48 trials, and each trial contained a unique target that appeared only once during the entire experiment. Each search display had one target and 24 distractors, and each object was placed in a random position on the screen. (This improves upon the display of Lupyan and Swingley (2012), who used a symmetrical grid layout for their search display. Having items displayed in completely random positions is more typical in visual search paradigms.)

Procedure. Participants were instructed to search for a target object among distractor objects. Both the target and the word to be repeated were displayed in the form of a word (as opposed to in the form of a picture). The target label and "repeat aloud" label remained on the screen until the participant pressed the "ENTER" key to begin the trial. This was followed by a screen instructing them to "keep repeating", which lasted four seconds. This was to ensure that the word was said at least a few times, in the event that the target was found right away. Then the search array appeared. Participants were instructed to press the spacebar when they found the target, as quickly as possible. RTs were measured from the onset of the search display to the spacebar press. After the spacebar response, the search array disappeared and four numbers appeared on the screen in various locations; one of the numbers appeared in the location of the target object. The participants were then given a choice between two numbers, both of which had appeared on the previous screen and one of which was the correct location of the target. The "F" and "J" keys were used to select the answer, to indicate that they did indeed correctly find the target during search (Figure 2). This procedure allowed for a better measure of RTs than that of Lupyan and Swingley (2012)—which had participants move the mouse to the target object on the screen and click on it—while still measuring accuracy. There were eight practice trials at the beginning of the experiment, two of each condition. A tape recorder and the physical presence of an experimenter ensured that participants continued speaking throughout every search. The entire experiment lasted approximately one hour, with a break given halfway.

Experiment 2

The purpose of Experiment 2 was to replicate Experiment 1 on a much larger display. The hypothesis was that the larger display would increase search times because it would physically take longer for the eyes to move the distance from one object to the next during search. This would potentially allow us to differentiate the results of the four conditions even further.

Participants. 38 new participants were recruited from the Arizona State

University Psychology 101 subject pool, and were given course credit for their participation. All were native English speakers and had normal or corrected-to-normal vision by self-report. Six subjects were excluded from data analysis based on the same performance criteria as in Experiment 1.

Apparatus. Stimuli were presented using Dell computers and 42-inch monitors, (i.e. big-screen TVs). This was the key difference between this experiment and Experiment 1. Participants responded via keyboard. As in Experiment 1, data was collected on up to two computers simultaneously, each sharing identical hardware and software profiles and each in its own room.

Stimuli, design, and procedure. Experiment 2 had the exact same stimuli, design and procedure as Experiment 1, with the sole difference being the size of the visual search display.

Experiment 3

The purpose of Experiment 3 was to apply eye tracking to our paradigm. Measuring eye movements allowed us to evaluate whether self-directed speech influences target locating (i.e. attentional guidance) or target identification after location (i.e. decision time).

Participants. 35 new participants were recruited from the Arizona State University Psychology 101 subject pool, and were given course credit for their participation. All were native English speakers and had normal or corrected-to-normal vision by self-report. Exclusion criteria was the same as in Experiments 1 and 2, with the addition of calibration measures – if the number of trials in which a participant's eye movements did not correctly calibrate was two or more standard deviations higher than the group mean, he or she was excluded from analysis. After these criteria, seven participants were excluded, resulting in an n of 28.

Apparatus. Data was collected one participant at a time using a Dell Optiplex 755 PC (2.66 GHz, 3.25 GB RAM). The monitor was 20 inches measured diagonally. E-Prime 2.0 software was used to control stimulus presentation and collect responses. Eye-movements were recorded by an Eyelink 1000 eye-tracker (SR Research Ltd., Mississauga, Ontario, Canada), mounted on the desktop. Participant vision was binocular, but only the left eye's movements were measured. A chinrest kept the viewing angle consistent across all participants while also stabilizing head movements to improve calibration.

Stimuli, design, and procedure. Experiment 3 had the exact same stimuli, design and procedure as Experiments 1 and 2, with the key addition of eye tracking. Participants were initially calibrated to ensure accurate tracking. If needed, participants were recalibrated during the experiment. Interest areas were defined by an invisible rectangular area that contained each given image. Unlike in Experiments 1 and 2, the trial procedure included a gaze-contingent fixation cross that appeared immediately before the search display onset. Participants had to maintain gaze on the cross for 500 ms (while still repeating the "repeat aloud" word), which triggered the search display to appear and ensured proper calibration. In rare instances where fixation was not achieved within 10 s, the trial was discarded and recalibration was performed. The entire experiment lasted a little over one hour, with a break given halfway.

RESULTS

Experiment 1

For accuracy and overall reaction times, repeated measures analyses of variance (ANOVAs) were performed with labeling condition as a within-subject factor, with Bonferroni adjustments for multiple comparisons. RT analyses were performed only on correct responses. Follow-up t-tests compared speaking conditions with each other. Accuracy and RT results are depicted in Figure 3.

Overall, participants were quite accurate across all four conditions (M = 96%). There was an overall effect of speaking condition, $F(3, 51) = 7.63$, $p < .001$, Cohen's d = .76. The Target condition (M = 98%) was significantly more accurate than the Nonword condition (M = 96%), $t(106) = 3.54$, $p = .001$, the Distractor Absent condition (M = 96%), $t(106) = 2.83$, $p = .006$, and the Distractor Present condition (M = 95%), $t(106) = 3.80$, $p < .001$. There were no significant differences in accuracy among the Nonword, Distractor Absent, and Distractor Present conditions, $ts < .96$.

The results for RTs showed similar patterns. There was an overall effect of condition, $F(3, 51) = 6.45$, $p < .001$, Cohen's d = .70. The Target condition (M = 2,406 ms) was significantly faster than the Nonword condition (M = 2,680 ms), $t(106) = 2.58$, $p = .011$, the Distractor Absent condition (M = 2,621 ms), $t(106) = 2.03$, $p = .045$, and the Distractor Present condition (M = 2,700 ms), $t(106) = 2.83$, $p = .006$. There were no significant differences in RTs among the Nonword, Distractor Absent, and Distractor Present conditions, $ts < .71$.

Experiment 2

Analyses used for Experiment 2 were the same as in Experiment 1. For the most part, results from Experiment 1 were replicated, simply on a larger scale (Figure 3). Again, participants were quite accurate across all four conditions ($M = 97\%$). However, there were only marginal differences in accuracy among the conditions, $F(3, 29) = 2.26$, $p = .087$.

The results for RTs showed similar patterns to those of Experiment 1. There was an overall effect of condition, $F(3, 29) = 13.13$, $p < .001$, Cohen's $d = 1.30$. The Target condition ($M = 2{,}623$ ms) was significantly faster than the Nonword condition ($M = 3{,}159$ ms), $t(62) = 3.57$, $p = .001$, the Distractor Absent condition ($M = 3{,}074$ ms), $t(62) = 3.35$, $p = .001$, and the Distractor Present condition ($M = 3{,}096$ ms), $t(62) = 3.67$, $p = .001$. Qualitatively, the mean differences between the Target condition and the other conditions were much larger than in Experiment 1, about 500 ms vs. 260 ms. As in Experiment 1, however, there were no significant differences in RTs among the Nonword, Distractor Absent, and Distractor Present conditions, *ts* < .54.

Experiment 3

Behavioral. Analyses used for Experiments 1 and 2 were used for the behavioral portion of Experiment 3. Results differed slightly from Experiments 1 and 2. Again, participants were quite accurate across all four conditions ($M = 98\%$). There was an overall effect of speaking condition on accuracy, $F(3, 25) = 3.84$, $p = .013$, Cohen's $d = .75$. The Target condition ($M = 99\%$) was significantly more accurate than the Nonword condition ($M = 97\%$), $t(54) = 2.99$, $p = .004$, the Distractor Absent condition ($M = 97\%$), *t*(54) = 2.75, *p* = .008, and the Distractor Present condition (*M* = 98%), *t*(54) = 2.48, *p* = .016. There were no significant differences in accuracy among the Nonword, Distractor Absent, and Distractor Present conditions, *ts* < .88.

The results for RTs showed patterns that differed slightly from those in Experiment 1. There was an overall effect of condition, $F(3, 25) = 4.74$, $p < .004$, Cohen's $d = .84$. The Target (*M* = 2,604 ms) and Nonword (*M* = 2,740 ms) conditions did not differ from each other, $t(54) = 0.95$, $p = .35$. RTs in the Target condition were faster than the Distractor Absent condition (*M* = 2,990 ms), $t(54) = 2.31$, $p = .025$, and were marginally faster than the Distractor Present condition (*M* = 2,873 ms), $t(54) = 1.93$, $p = .059$. As in Experiment 1, however, there were no significant differences in RTs among the Nonword, Distractor Absent, and Distractor Present conditions, *ts* < 1.4.

Eye Movements. ANOVAs and follow-up t-tests were used to examine the following eye movement measures: overall RTs; accuracy (Figure 3); the amount of time it takes for the eyes to fixate on the target (i.e. time to target fixation); the amount of time between target fixation and keyboard response (i.e. decision time); the number of distinct distractors fixated before the target; the number of times the eyes refixated on a distractor; target, distractor, and foil (foil applies to Distractor Present condition only) fixation duration (Figure 4); and sequential analyses of trials in the Distractor Present condition (Figure 5). Only accurate trials in which the target was eventually fixated were included in analyses.

RTs and accuracy. The first and most important distinction made possible with eye tracking is the ability to determine whether or not a foil (the object that is repeated aloud but not searched for) was fixated during Distractor Present (DP) trials. This was important because if a foil was present on the screen but was never fixated, then theoretically that trial becomes functionally equivalent to the Distractor Absent trials. The overall RT differences become much more pronounced when the DP condition only contains trials where foils were fixated, $F(3, 25) = 54.118$, $p < .0001$, Cohen's $d = 2.83$. In Experiments 1 and 2, means for DP condition RTs were 2,700 ms and 3,096 ms, respectively. Experiment 3 had similar results ($M = 2{,}873$ ms), but when you only look at the trials in which the foil was fixated, RTs increase to 4,474 ms. Not surprisingly, RTs for these trials were significantly slower than in the Target condition, $t(54) = 7.49$, $p < .0001$, the Nonword condition, $t(54) = 6.79$, $p < .0001$, and the Distractor Absent condition, $t(54) = 5.50$, $p < .001$. Accuracy in trials where the foil was fixated also decreased ($M = 96\%$), and was significantly lower than accuracy in the Target condition, $t(54) = 3.30$, $p < .002$, but not the Nonword or Distractor Absent conditions (*ts* = 1.5).

The foil was fixated in 30% of trials in the DP condition. In the remaining 70% of trials, in which the foil was not fixated, RTs decreased dramatically ($M = 2{,}122$) (see Figure 3). Participants were substantially faster in these 70% of DP trials than they were in the trials where the foil was fixated, $t(54) = 9.73$, $p < .0001$. They were also faster than in all of the other conditions – Target, $t(54) = 4.13$, $p < .001$, Nonword, $t(54) = 4.80$, $p < .001$, and Distractor Absent, $t(54) = 5.596$, $p < .001$. They were also significantly more accurate ($M = 99\%$) than they were in foil-fixated trials, $t(54) = 3.38$, $p = .001$, and were additionally more accurate than in the Nonword condition, $t(54) = 3.08$, $p = .003$, and the Distractor Absent condition, $t(54) = 2.86$, $p = .006$. In short, in the 70% of DP trials where the foil was not fixated, performance drastically increased. Speculation as to why this effect occurred is considered in the Discussion section, but nevertheless this effect is seen in every single informative eye movement measure we assessed, and thus will be included throughout the Results section.

Time to target fixation. After the onset of the search display, the time to target fixation showed a similar overall pattern of results as that of RTs, with participants being fastest to locate the target in the DP trials where the foil was not fixated, (M = 1,378 ms), followed by the Target condition (M = 1,703 ms), Nonword condition (M = 1,775 ms), Distractor Absent Condition (M = 1,851 ms), and the DP trials where the foil *was* fixated (M = 2,747 ms). There was an overall effect of condition[1], $F(3, 25) = 48.14$, $p < .001$, Cohen's $d = 2.67$. The Target condition did not differ from the Nonword condition, $t(54) = 0.72$, $p = .48$, or the Distractor Absent condition, $t(54) = 1.46$, $p = .15$, and the Nonword and Distractor Absent conditions did not differ from each other, $t(54) = 0.66$, $p = .51$. The DP condition differed significantly from all other conditions, both in trials where the foil was fixated and in trials where it was not. In trials where the foil was fixated, participants were much slower to fixate the target than in the Target condition, $t(54) = 7.15$, $p < .001$, the Nonword condition, $t(54) = 6.26$, $p < .001$, and the Distractor Absent condition, $t(54) = 5.74$, $p < .001$. In DP trials where the foil was *not* fixated, participants were much faster to fixate the target than in foil-fixated trials, $t(54) = 9.57$, $p < .0001$. These trials were also faster than the Target condition, $t(54) = 4.03$, $p < .001$, the

[1] For consistency across analyses, all eye movement ANOVAs contained the four conditions (Target, Nonword, Distractor Absent, and Distractor Present), where the DP condition only contained trials in which the foil was fixated. While DP trials in which the foil was *not* fixated are included in t-tests and are scientifically interesting, they were not included in ANOVAs as part of the DP condition or as their own condition as they were not central to the theoretical questions, original hypotheses, or overall paradigm of this project.

Nonword condition, $t(54) = 4.12$, $p < .001$, and the Distractor Absent condition, $t(54) = 4.85$, $p < .001$.

Decision time. The time in between target fixation and the keyboard press again showed the same general patterns as time-to-target-fixation and overall RTs. Participants were fastest to make a keyboard press in the DP trials where the foil was not fixated, ($M = 744$ ms), followed by the Target condition ($M = 901$ ms), Nonword condition ($M = 965$ ms), Distractor Absent Condition ($M = 1,139$ ms), and the DP trials where the foil *was* fixated ($M = 1,727$ ms). There was an overall effect of condition, $F(3, 25) = 30.17$, $p < .001$, Cohen's $d = 2.12$. The Target condition did not differ from the Nonword condition, $t(54) = 0.83$, $p = .41$, and the Nonword and Distractor Absent conditions did not quite differ from each other, $t(54) = 1.78$, $p = .08$, but decision times in the Target condition were faster than the Distractor Absent condition, $t(54) = 2.49$, $p = .016$. The DP condition again differed significantly from all other conditions, both in trials where the foil was fixated and in trials where it was not. In trials where the foil was fixated, participants had slower decision times than in the Target condition, $t(54) = 6.18$, $p < .001$, the Nonword condition, $t(54) = 5.63$, $p < .001$, and the Distractor Absent condition, $t(54) = 4.00$, $p < .001$. In DP trials where the foil was *not* fixated, decision times were much faster than in foil-fixated trials, $t(54) = 7.55$, $p < .001$. These trials were also faster than the Target condition, $t(54) = 2.35$, $p = .022$, the Nonword condition, $t(54) = 3.15$, $p = .003$, and the Distractor Absent condition, $t(54) = 4.63$, $p < .001$.

No. of distractors fixated. The number of distinct distractors fixated before target fixation differed depending on condition, $F(3, 25) = 24.35$, $p < .001$, Cohen's $d = 1.90$.

Significant changes in the number of distractors fixated in a trial explains the differences in the total duration of distractor fixations that occurred before target fixation. On average, participants looked at nearly two more distractors in DP foil-fixated trials (M = 7.32) than they did in the Target condition (M = 5.61), $t(54) = 5.34$, $p < .001$, the Nonword condition (M = 5.73), $t(54) = 4.97$, $p < .001$, or the Distractor Absent condition (M = 5.82), $t(54) = 4.47$, $p < .001$. Participants looked at fewer distractors in DP trials where the foil was *not* fixated (M = 4.68) than they did in DP foil-fixated trials, $t(54)$ = 8.84, $p < .001$. The number of distractors fixated in DP trials where there was no foil fixation was also significantly lower than that of the Target condition, $t(54) = 4.50$, $p <$ .001, the Nonword condition, $t(54) = 5.07$, $p < .001$, and the Distractor Absent condition, $t(54) = 4.93$ $p < .001$. There were no significant differences in the number of distractors fixated among the Target, Nonword, or Distractor Absent conditions, $ts < .80$.

No. of distractor refixations. The number of distractor refixations (i.e. fixating on an item, looking away, and then refixating on the item) differed depending on condition, $F(3, 25) = 10.19$, $p < .001$, Cohen's $d = 1.23$. On average, participants refixated items more frequently in DP foil-fixated trials (M = 1.42) than they did in the Target condition (M = .78), $t(54) = 3.58$, $p = .001$, the Nonword condition (M = .86), $t(54) = 3.16$, $p = .003$, or the Distractor Absent condition (M = .94), $t(54) = 2.56$, p = .013. Participants made fewer refixations in DP trials where the foil was *not* fixated (M = .46) than they did in DP foil-fixated trials, $t(54) = 5.74$, $p < .001$. The number of refixations in DP trials where there was no foil fixation was also significantly lower than that of the Target condition, $t(54) = 3.44$, $p < .001$, the Nonword condition, $t(54) = 4.63$,

$p < .001$, and the Distractor Absent condition, $t(54) = 4.53$ $p < .001$. There were no significant differences in number of refixations among the Target, Nonword, or Distractor Absent conditions, $ts < 1.3$.

Fixation durations.[2] The duration of target fixation only varied marginally by condition, $F(3, 25) = 2.51$, $p = .065$, Cohen's $d = .61$. DP foil-fixated target fixation durations ($M = 359$ ms) were marginally shorter than those in the Target condition ($M = 392$ ms), $t(54) = 1.64$, $p = .11$, and the Distractor Absent condition ($M = 389$ ms), $t(54) = 1.70$, $p = .10$. Participants did have significantly longer target fixations in DP no-foil-fixation trials ($M = 404$ ms) than in DP trials where the foil *was* fixated, $t(54) = 2.40$, $p = .02$. There were no significant differences among any other conditions, $ts < 1.2$.

The durations of distractor (and subsequently foil) fixations[3] differed by condition, $F(4, 24) = 5.16$, $p = .001$, Cohen's $d = .87$. Distractor fixations in DP foil-fixated trials ($M = 210$ ms) were only marginally longer than trials in which the foil was not fixated ($M = 199$ ms), $t(54) = 1.64$, $p = .11$, or the Target condition ($M = 199$ ms), $t(54) = 1.66$, $p = .10$. Fixations were longest on the foil ($M = 221$ ms), and were significantly different from distractor fixations in the Target condition, $t(54) = 2.15$, $p = .036$, the Distractor Absent condition ($M = 199$ ms), $t(54) = 2.02$, $p = .048$, or in DP trials where the foil was not fixated, $t(54) = 2.14$, $p = .037$. Foil fixations were marginally longer than distractor fixations in the Nonword condition ($M = 200$ ms), $t(54) = 1.96$, $p = .055$. No other conditions differed from each other, $ts < 1.4$. In the Distractor Present

[2] Unless otherwise specified, fixation durations are first-run, meaning refixations on items are not included in the analysis.
[3] In addition to including distractor fixations in the Target, Nonword, Distractor Absent, and Distractor Present foil-fixated conditions, fixations on the actual foils themselves were included as a fifth measure, as foils are technically distractors.

condition, the duration of distractor fixations within a single trial did not change after a foil was fixated, $t(54) = 0.19$, $p = .85$.

The total duration of all pre-target distractor fixations, including item refixations, differed significantly by condition, $F(3, 25) = 24.31$, $p < .001$, Cohen's $d = 1.90$. Participants spent more time looking at distractors in DP foil-fixated trials ($M = 1{,}985$ ms) than they did in the Target condition ($M = 1{,}363$ ms), $t(54) = 5.36$, $p < .001$, the Nonword condition ($M = 1{,}439$ ms), $t(54) = 4.37$, $p < .001$, or the Distractor Absent condition ($M = 1{,}471$ ms), $t(54) = 4.12$, $p < .001$. Participants spent less time fixating on distractors in DP trials where the foil was *not* fixated ($M = 1{,}087$ ms) than they did in DP foil-fixated trials, $t(54) = 7.91$, $p < .001$. Total pre-target distractor fixation duration in DP trials where there was no foil fixation was also significantly less than that of the Target condition, $t(54) = 3.99$, $p < .001$, the Nonword condition, $t(54) = 4.23$, $p < .001$, and the Distractor Absent condition, $t(54) = 4.64$, $p < .001$. There were no significant differences among the Target, Nonword, or Distractor Absent conditions, $ts < 1.2$.

DP sequential analyses. Unplanned sequential trial analyses were performed in an attempt to better understand the effect observed in the Distractor Present condition, wherein performance dramatically increased in trials where the foil was not fixated. We performed pairwise comparisons on overall RTs in trials in the DP condition, including: 1) trials where the foil was fixated, 2) the first trial immediately following a foil-fixation trials, and 3) the second trial following foil-fixation trials. In trials immediately following a trial in which the foil was fixated, RTs quickened significantly, by 1,454 ms, $t(54) = 5.32$, $p < .001$. RTs did not change significantly further in the following trial (i.e. two trials after a foil fixation), $t(54) = 0.41$, $p = .68$, even though they did quicken by 82 ms. The approximately 1,500 ms described here, while statistically significant, alone is not enough to explain the 2,352 ms difference between DP trials with and without foil fixation. Figure 5 depicts each trial for two participants in the DP condition, to show the overall trends that occur around a foil fixation.

DISCUSSION

This work utilized both RTs and eye movements to examine the Label Feedback Effect through four simple visual search tasks. The aim was to 1) evaluate whether self-directed speech influences attentional guidance or decision time, and 2) determine whether the Label Feedback Effect reflects changes in visual attention, or some other mechanism (e.g. template maintenance in working memory). Participants simultaneously spoke aloud while performing visual search, repeating either target names, nonwords, the names of real objects not in the display, or the names of distractor objects that were present on-screen. Measuring eye movements aided in the exploration of the Label Feedback Hypothesis in ways that RTs alone could not.

We found that speaking target names weakly facilitates visual search, but speaking different names strongly inhibits search. Eye tracking gave us insight into the exact mechanisms behind the effect, the most important of which are the time until target fixation and decision time, which combine together to form the overall RTs. Interestingly, both attentional guidance *and* decision time contribute seemingly equally to the effect. Repeating target names had little effect on attentional guidance or perceptual decisions, with results equivalent to the Nonword condition. In fact, differences among the Target and Nonword conditions were not significant in every single eye movement measure, which suggests that the effect does not result from top-down influences of language. Because the Nonword condition is devoid of linguistic content, if the effect was purely a language phenomenon then we would expect to see differences between the Target and Nonword conditions. Additionally, slower attentional guidance and decision time in the Distractor Present (foil-fixated) condition suggest that language inhibits a person's ability to perform the visual search task in a strategically efficient way. Speaking distractor names had powerful effects on guidance and decision-making, which was especially clear in DP trials, when people actually fixated the named distractor. In those trials, it took participants a full extra second for their eye to land on the target. This is because they fixated nearly two additional distractors, made additional refixations, fixated longer on the foil, and in total spent nearly an extra 500 ms fixating distractors before finally fixating the target. When the eyes did eventually find the target, it took nearly a full extra second to initiate a response. Overall, repeating target names may help people maintain search templates, avoiding capture by background objects, whereas speaking unrelated (or intentionally misleading) names does the opposite. The most parsimonious account, therefore, is that language affects target maintenance during search, rather than visual perception.

This research significantly improves upon previous research in the area, and gives us insight into the mechanisms behind the Label Feedback Effect. In all three experiments, our RT effects were significantly larger than that of Lupyan and Swingley (2012) (approximately 50 ms vs. 300-2300 ms), and our accuracy, while still high, was not as ceilinged as previous research and is therefore potentially more informative. The results of Lupyan and Swingley's (2012) research are difficult to interpret, not least of all because silently reading a word is an imperfect control. It is comparing a language task to a question mark – it is difficult to know the true mechanisms at work. The Nonword condition in the current project, however, provided consistency through a dual-task where true linguistic content was removed. Additionally, the current work's overall paradigm is more in keeping with traditional visual search paradigms, utilizing keyboard press for more accurate RTs and a random non-grid item display for more realistic search.

While the eye movement data is undeniably more informative, we acknowledge that the behavioral results of Experiments 1 and 2 vs. Experiment 3 are somewhat inconsistent. In Experiments 1 and 2 the Target condition is statistically different from all the other conditions in both RTs and accuracy, which potentially lends itself to Lupyan and Swingley's (2012) conclusions that "self-directed speech activated visual properties of the target category over and above silently reading the word." Unlike the first two experiments, Experiment 3 depicted a differentiation by conditions, which is what we originally predicted. The most significant change appears to take place in the relationship of the Nonword condition to the other conditions (see Figure 3). This discrepancy between experiments held true when eye movement filters (e.g. only including trials in which the target was fixated or DP trials where the foil was fixated) were removed. There is seemingly nothing to explain this discrepancy, as testing conditions were quite consistent across all three experiments. Sample size did differ ($n_1 = 54$, $n_2 = 32$, $n_3 = 28$), but observed power was quite high for all experiments (for RTs: $power_1 = .97$, $power_2 = .99$, and $power_3 = .89$). Different monitor sizes, while affecting the scale of RTs, should not affect the overall relationship of the four conditions to one another. Because participants had to be calibrated on the eye tracker, Experiment 3 was about 10 to 15 minutes longer than Experiments 1 or 2, and while it is feasible that participants in Experiment 3 were more fatigued, blocks were randomized and there is therefore no reason for the overall pattern—or specifically the relationship of the Nonword condition—to have changed. In Experiment 3, the researcher sat right next to the participants, whereas in Experiments 1 and 2 the researcher sat just outside the room with the door cracked, but those participants were made aware that the researchers were monitoring their speaking and that their voices were being recorded. If there was an overall motivation effect across experiments, we would have expected RTs to be faster in Experiment 3, but in fact they are very similar to Experiment 1 (RTs in Experiment 2 are understandably not comparable due to the larger search display). And again this should not change the overall relationship of the four conditions to one another. The only other difference across experiments is that participants in Experiment 3 used a chin rest, and so were kept at a consistent viewing angle and were more constrained to do the task the way it was intended. This is a potential argument for the superiority of the Experiment 3 data.

In the Distractor Present condition, eye movements revealed that the foil was fixated in 30% of trials. In the remaining 70% of trials, in which the foil was not fixated, something curious happened. RTs decreased dramatically, by more than 2,300 ms. Accuracy improved from 96% to 99%. In short, performance drastically increased, not just compared to the DP trials where the foil *was* fixated, but to trials in all other conditions, as well. One possible explanation for this would be a statistical selection effect. There are 25 items in the search array, and on average participants looked at 5 to 7 distractors before fixating the target, which means that approximately 25% of the objects on the screen are fixated in any given trial. In this account, about one third of the time a person's eyes happen to wander into the region of the screen where the foil is located, resulting in a lapse in ability to perform the task. In the other 70% of the trials, a person's eyes never approach the foil, and on average they only fixate approximately four items. This means that mathematically that a person only really looks at 75% of the screen, and therefore just happens to find the target sooner. This would potentially explain the observed effect, however, the fact that decision time is also affected (instead of only time to target fixation) makes this explanation alone unlikely, because decision time should not be influenced by chance screen layout.

Another possible explanation for this phenomenon is the Gratton Effect (Gratton, Coles, and Donchin, 1992). The Gratton Effect was originally observed in the flanker task, where the flanker effect was smaller (i.e. performance increased) immediately after incongruent trials than after congruent trials. The standard explanation for the Gratton Effect is that, after an incongruent trial, relatively more attention is directed to relevant task information while better filtering out the irrelevant information, reducing the congruency effect. This effect has also been demonstrated in several other areas of cognitive psychology (e.g. the Stroop effect, Kerns et al., 2004). In the present study, it would suggest that after a trial in which the participant sees the foil, she recalibrates, her attention refocuses, and she has a re-honed decision strategy which then carries over to subsequent trials. While this explanation seems to fit, it might be presumptuous, as there are many crucial established components to the Gratton Effect that we cannot clearly identify in our paradigm. Firstly, we cannot pinpoint what does and does not constitute an "incongruent" or difficult trial (and that definition may very well change with individual differences). Is a difficult trial one in which the foil is fixated? This would make intuitive sense, but then we would expect to see a dramatic decrease in the RTs of the immediately subsequent trials. We do see a decrease, and it is a large one (about 1500 ms), but it is not large enough to account for the average RT of 2116 ms on Distractor Present trials where the foil is not fixated. Perhaps a trial can also be considered difficult if the participant "sees" the foil in her peripheral vision, allocating attention and effort to *not look at it*—this would not be revealed by a foil fixation, and so is difficult to measure. Additionally, what constitutes a "correct response", within the context of a typical Gratton Effect (e.g. from a Flanker or Stroop task)? Participants were nearly at ceiling in accuracy, with several participants making zero errors during the Distractor Present trials. Is it only an error if the participant doesn't ultimately find the correct target? One might also consider the very act of fixating on the foil an error, given that the participant is not supposed to be looking at the thing she is repeating aloud. Taken all together, it seems a bit hasty to classify this perplexing effect as a Gratton Effect, though it would best explain the observed effect and should not be ruled out as a possibility. Of course, it could potentially be both a statistical selection effect combined with a Gratton-type carryover effect. Future projects could examine this observed phenomenon closer.

Additional questions involve examining precisely what factors moderate target maintenance and attention in the current paradigm – perhaps the four conditions have a different effect for people with varying levels of working memory, wherein saying the target name is more helpful and saying the foil name more costly for people with low working memory. Overall, the present paradigm offers a collection of manipulations that allowed us to explore both verbal and attention processes, separately and together, to better illuminate the extent to which speech impacts perception in visual search. This work examines the Label Feedback Hypothesis in a novel way, and eye movement measures provide a clearer picture in which language affects target maintenance during search, but not visual perception. The question of whether the top-down influences of language shape low-level visual and auditory perception is a controversial one. In the current study, while language does impact performance on nonlinguistic visual search tasks, evidence suggests that over and above *language's* (i.e. linguistic meaning) effect on performance, the effect can better be explained by the attentional processes accompanied by it.

REFERENCES

Boutonnet, B., and Lupyan, G. (2015). Words Jump-Start Vision: A Label Advantage in Object Recognition. *The Journal of Neuroscience*, *35*(25).

Brady, T. F., Konkle, T., Alvarez, G. A. and Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences, USA*, 105 (38), 14325-14329.

Bruner, J. S., Goodnow, J. J., & Austin, G. A. (1956). A study of thinking. New York: Wiley.

Davelaar, E. J., & Stevens, J. (2009). Sequential dependencies in the Eriksen flanker task: A direct comparison of two competing accounts. *Psychonomic Bulletin & Review*, 16, 121- 126.

Deutscher, G. (2010). *Through the language glass: Why the world looks different in other languages*. Macmillan.

Drivonikou, G. V., Kay, P., Regier, T., Ivry, R. B., Gilbert, A. L., Franklin, A., et al. (2007). Further evidence that Whorfian effects are stronger in the right visual field than the left. *Proceedings of the National Academy of Sciences of the United States of America*, *104*, 1097–1102.

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

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.

Goldinger, S.D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, 105, 251-279.

Goldstone, R. L. (1994). Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology: General, 123*(2), 178-200.

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

Gratton, G., Coles, M. G. H., & Donchin, E. (1992). Optimizing the use of information – Strategic control of activation of responses. Journal of Experimental Psychology General, 121, 480–506.

Harnad, S. (1987). Categorical perception. Cambridge, England: Cambridge University Press.

Hout, M. C., Goldinger, S. D., & Brady, K. J. (2014). MM-MDS: A multidimensional scaling database with similarity ratings for 240 object categories from the *Massive Memory* picture database. *PLoS ONE, 9*: e112644. doi: 10.1371/journal.pone.0112644.

Kerns, J. G., Cohen, J. D., MacDonald, A. W., 3rd, Cho, R. Y., Stenger, V. A., & Carter, C. S. (2004, February 13). Anterior cingulate conflict monitoring and adjustments in control. Science, 303, 1023–1026

Klemfuss, N., Prinzmetal, W., & Ivry, R. B. (2012). How Does Language Change Perception: A Cautionary Note. *Frontiers in Psychology*, *3*, 78.

Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010). Conceptual distinctiveness supports detailed visual long-term memory for real-world objects. *Journal of Experimental Psychology: General*, 139(3), 558-78.

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

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.

Lupyan, G., & Spivey, M. J. (2010a). Redundant spoken labels facilitate perception of multiple items. *Attention, Perception & Psychophysics*, *72*(8), 2236–2253.

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

Lupyan, G., and Swingley, D. (2012). Self-directed speech affect visual search performance. *Quarterly Journal of Experimental Psychology*, i*First*, 1-18.

Lupyan, G., and Ward, E. J. (2013) Language can boost otherwise unseen objects into visual awareness. *Proceedings of the National Academy of Sciences. 110*(35) 1419-201.

Meteyard L, Bahrami B, Vigliocco G (2007). Motion detection and motion verbs: language affects low-level visual perception. *Psychological Science, 18*:1007–1013.

Regier, T., & Kay, P. (2009). Language, thought, and color: Whorf was half right. *Trends in Cognitive Sciences*, *13*(10), 439-446.

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.

Schneider, W., Eschman, A., & Zuccolotto, A. (2012). *E-Prime user's guide.* Pittsburgh, PA: Psychology Software Tools.

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* (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.

Suegami, T., Aminihajibashi, S., & Laeng, B. (2014). Another look at category effects on colour perception and their left hemispheric lateralisation: no evidence from a colour identification task. *Cognitive Processes.* 15:217–226

Whorf, B. (1956). Language, thought, and reality: Selected writings of Benjamin Lee Whorf. Ed. J. B. Carroll. *Cambridge: MIT Press*.

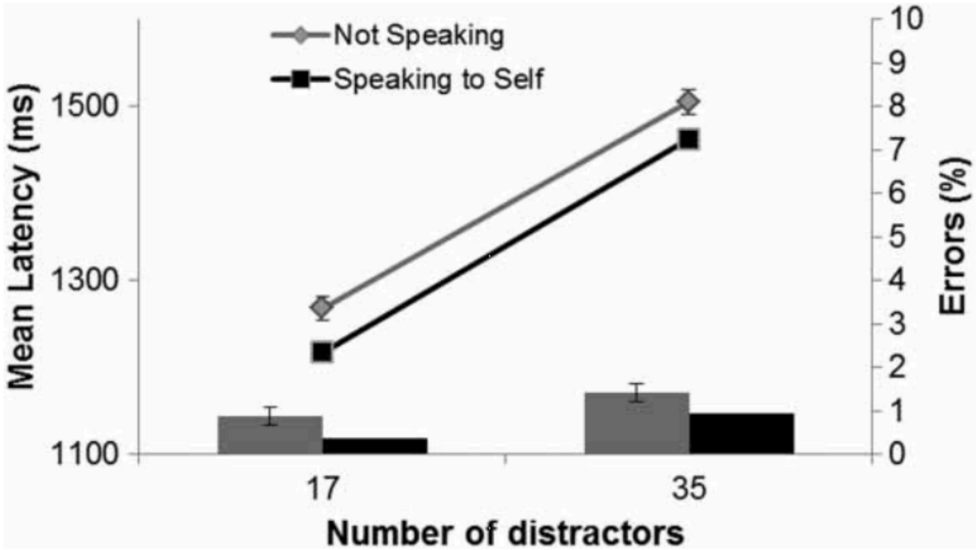


Figure 1. RTs (lines) and accuracy (bars) observed by Lupyan and Swingley (2012).

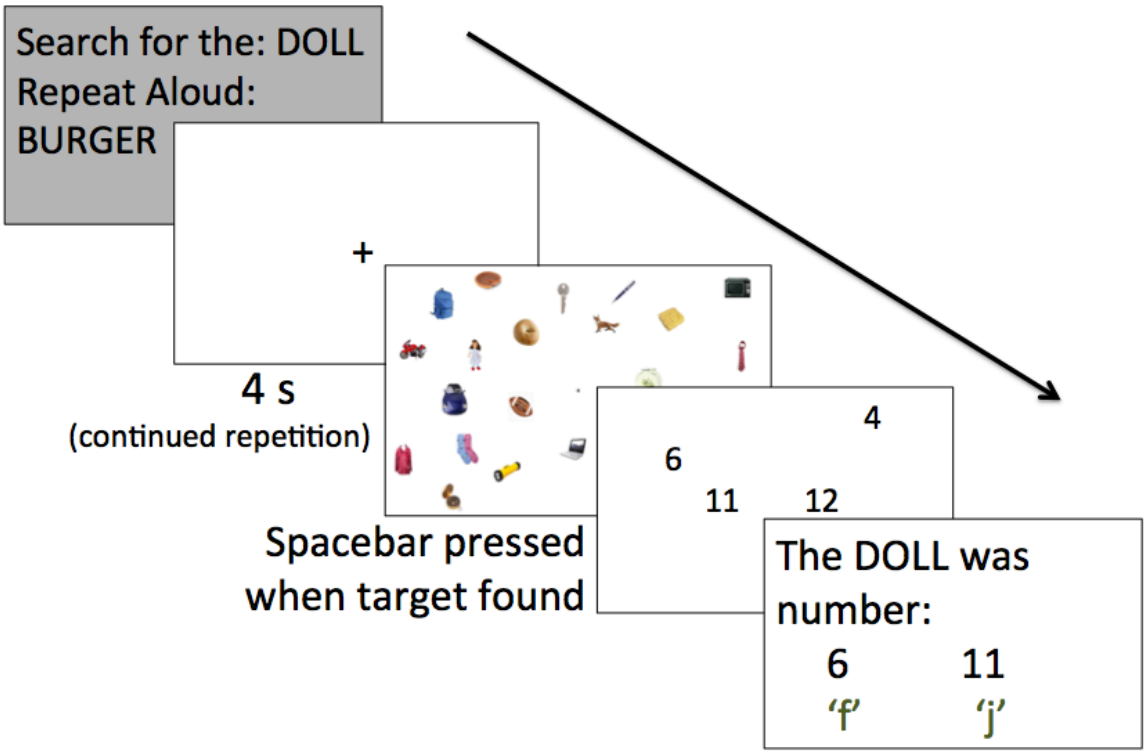


Figure 2. Experimental procedure, which was used in all three experiments.

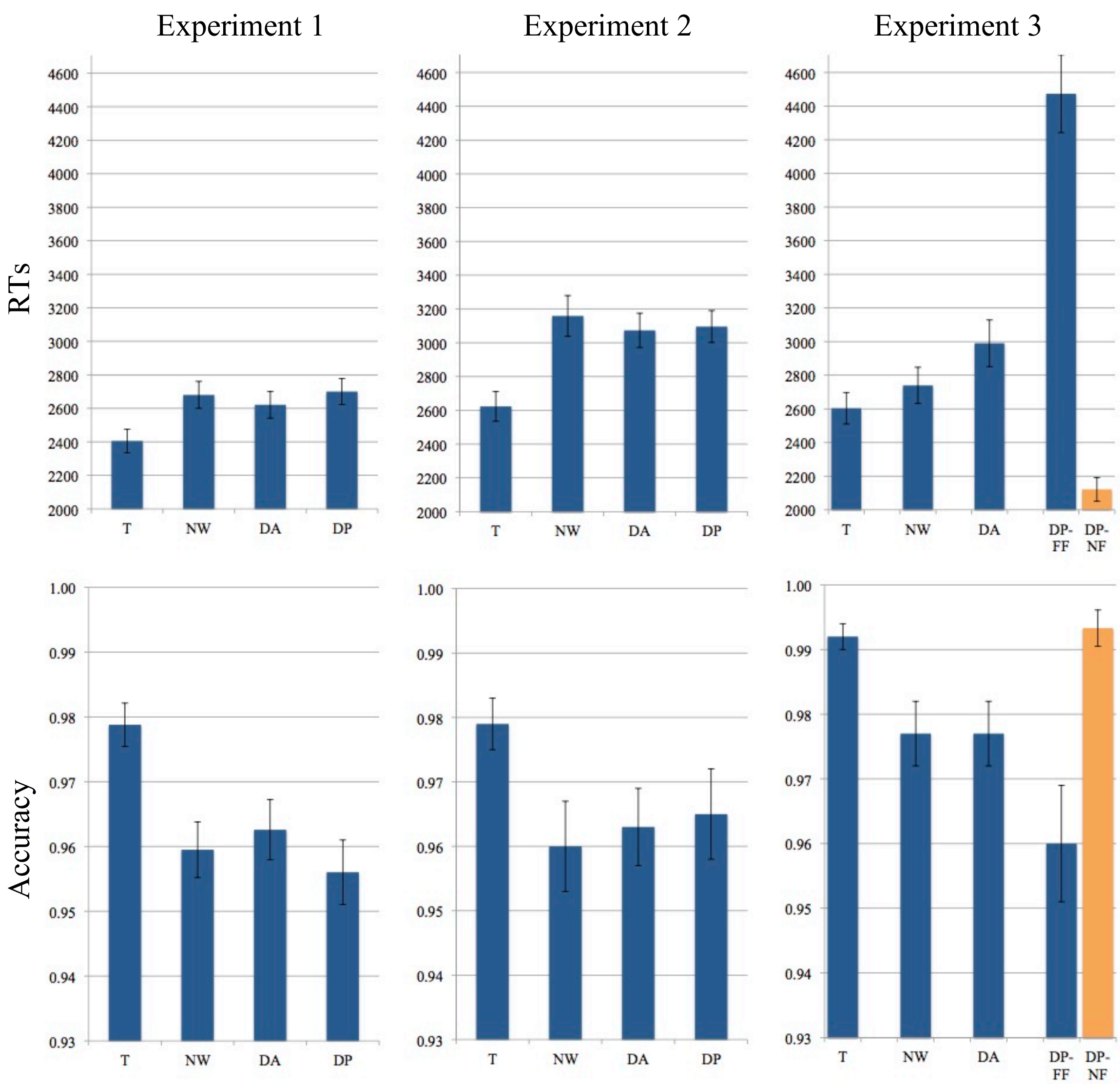
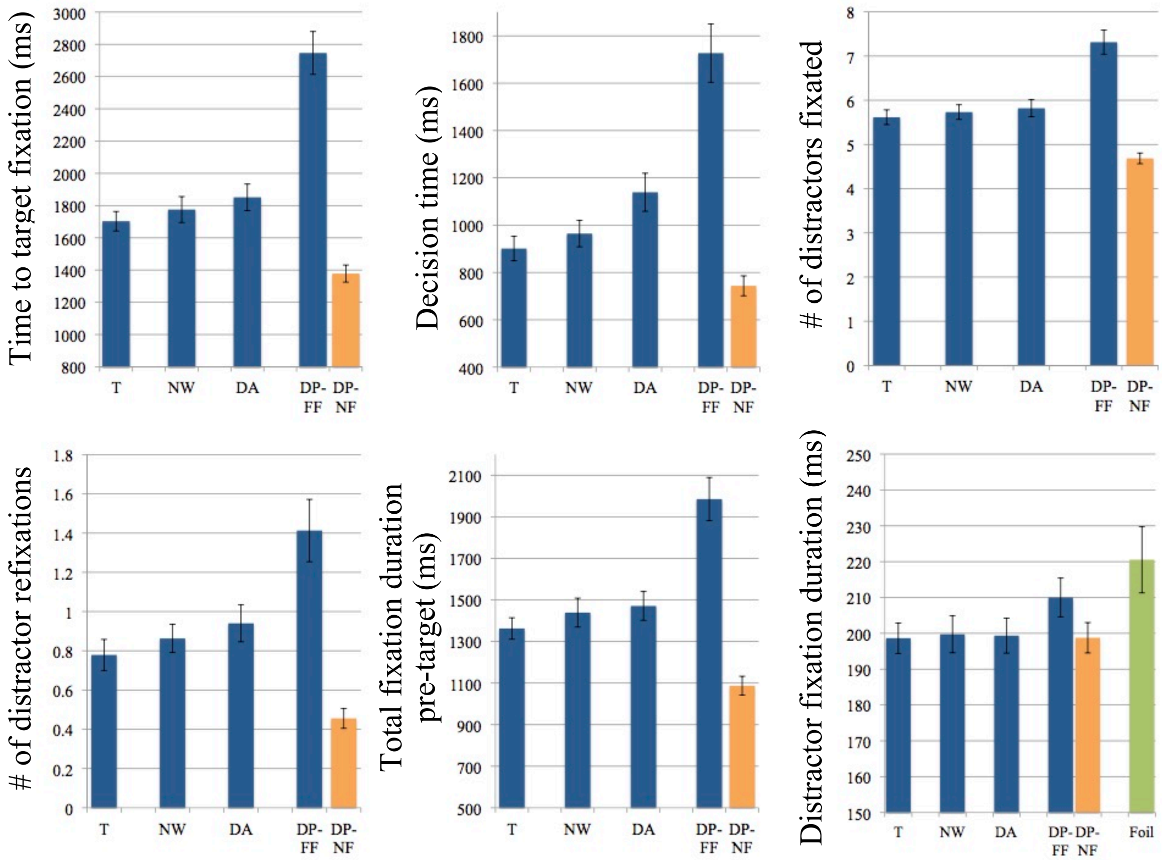


Figure 3. Experimental search reaction times (RTs) and accuracy for all three experiments. For every graph each applicable condition is shown: T = Target; NW = Nonword; DA = Distractor Absent; DP = Distractor Present (all trials); DP-FF = DP trials in which the foil was fixated; DP-NF = trials in which the foil was not fixated. Error bars represent ±1 standard error of the mean.



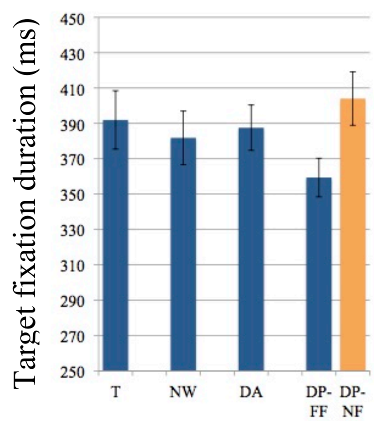


Figure 4. Eye tracking measures for Experiment 3. For every graph each applicable condition is shown: T = Target; NW = Nonword; DA = Distractor Absent; DP = Distractor Present (all trials); DP-FF = DP trials in which the foil was fixated; DP-NF = trials in which the foil was not fixated; Foil = the object that is repeated aloud but not searched for in DP trials. Error bars represent ±1 standard error of the mean.

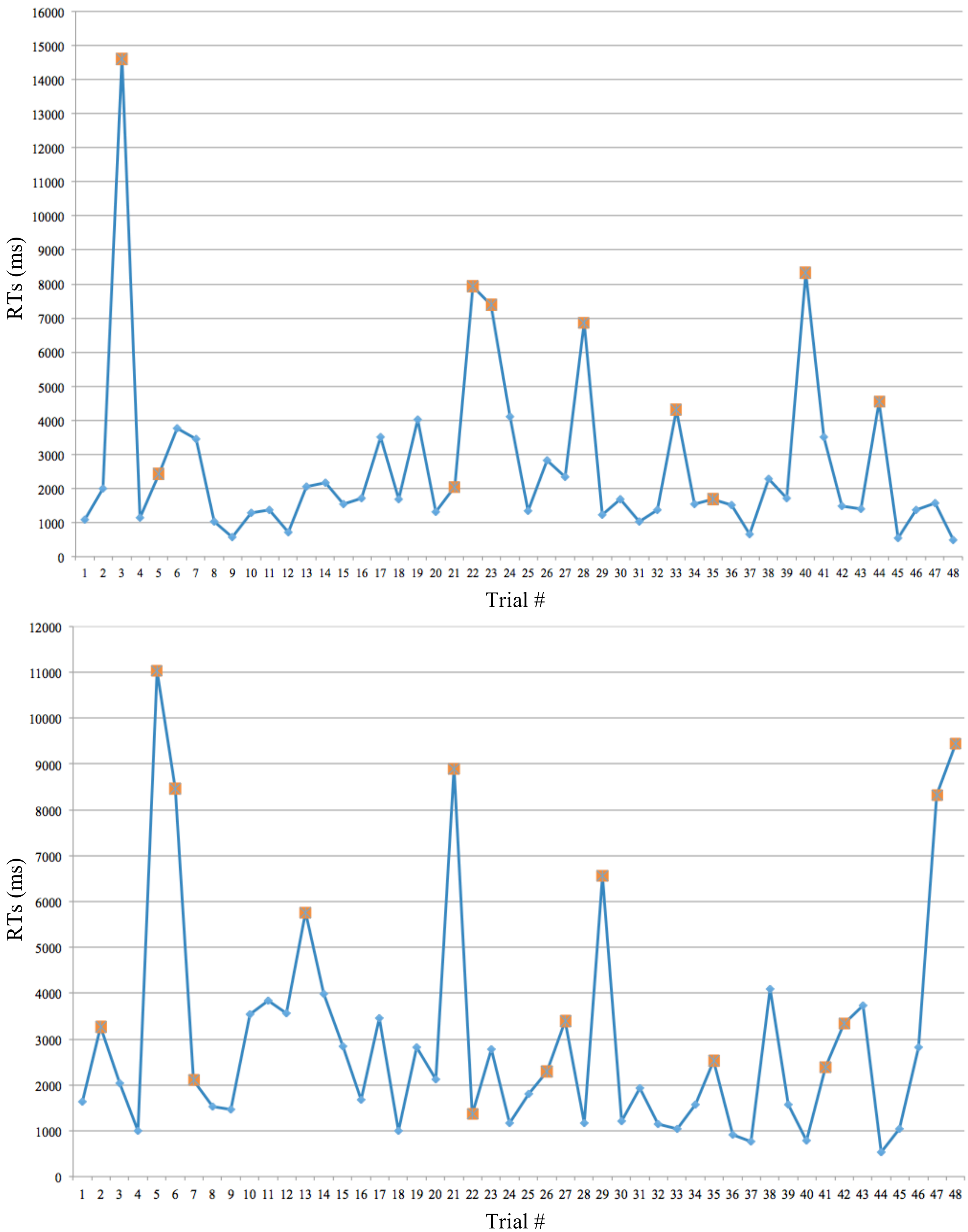


Figure 5. RTs for each trial for two participants in the Distractor Present condition, to show the overall trends that occur around a foil fixation. Each participant depicted had 100% accuracy.
= trials in which the foil was fixated.

APPENDIX A

IRB APPROVAL DOCUMENT

EXEMPTION GRANTED

Stephen Goldinger
Psychology
480/965-0127
goldinger@asu.edu

Dear Stephen Goldinger:

On 1/23/2015 the ASU IRB reviewed the following protocol:

Type of Review:	Initial Study
Title:	Language and Perception Research
Investigator:	Stephen Goldinger
IRB ID:	STUDY00002135
Funding:	None
Grant Title:	None
Grant ID:	None
Documents Reviewed:	• HRP-503a - Language and Perception Research PROTOCOL SOCIAL BEHAVIORAL-3.docx, Category: IRB Protocol; • Language and Perception _ Eye Tracker Explanation.pdf, Category: Participant materials (specific directions for them); • Language and Perceptiopn Recruitment Script-2.pdf, Category: Recruitment Materials; • LanguageAndPerception_coverletter-3.pdf, Category: Consent Form;

The IRB determined that the protocol is considered exempt pursuant to Federal Regulations 45CFR46 (2) Tests, surveys, interviews, or observation on 1/23/2015.

In conducting this protocol you are required to follow the requirements listed in the INVESTIGATOR MANUAL (HRP-103).

Sincerely,

IRB Administrator
cc: Katherine Jones
Stephen Goldinger