Differential Eye Movements in Verbal and Nonverbal Search

Robert S. Hurley[1,2]*, Jonathan Sander[1], Kayleigh Nemeth[1], Brittany R. Lapin[3], Wei Huang[2] and Mustafa Seckin[2]

[1]Department of Psychology, Cleveland State University, Cleveland, OH, United States, [2]Mesulam Center for Cognitive Neurology and Alzheimer's Disease, Northwestern University, Chicago, IL, United States, [3]Department of Quantitative Health Sciences, Cleveland Clinic, Cleveland, OH, United States

In addition to "nonverbal search" for objects, modern life also necessitates "verbal search" for written words in variable configurations. We know less about how we locate words in novel spatial arrangements, as occurs on websites and menus, than when words are located in passages. In this study we leveraged eye tracking technology to examine the hypothesis that objects are simultaneously screened in parallel while words can only be found when each are directly foveated in serial fashion. Participants were provided with a cue (e.g. rabbit) and tasked with finding a thematically-related target (e.g. carrot) embedded within an array including a dozen distractors. The cues and arrays were comprised of object pictures on nonverbal trials, and of written words on verbal trials. In keeping with the well-established "picture superiority effect," picture targets were identified more rapidly than word targets. Eye movement analysis showed that picture superiority was promoted by parallel viewing of objects, while words were viewed serially. Different factors influenced performance in each stimulus modality; lexical characteristics such as word frequency modulated viewing times during verbal search, while taxonomic category affected viewing times during nonverbal search. In addition to within-platform task conditions, performance was examined in cross-platform conditions where picture cues were followed by word arrays, and vice versa. Although taxonomically-related words did not capture gaze on verbal trials, they were viewed disproportionately when preceded by cross-platform picture cues. Our findings suggest that verbal and nonverbal search are associated with qualitatively different search strategies and forms of distraction, and cross-platform search incorporates characteristics of both.

Keywords: visual search, eye movements, language, nonverbal processing, picture superiority

INTRODUCTION

Humans regularly engage in visual search to navigate our cluttered world. We engage in "nonverbal search" for objects throughout the day, for example by picking out clothes in the morning, finding a parking space at work, or gathering ingredients to prepare dinner. Whereas many aspects of nonverbal search are shared with other animal species (Bichot et al., 2005; Goto et al., 2014), humans are unique in having written language, and we are frequently required to engage in "verbal search" through complicated arrays of words. A typical day may involve finding stories to read on a newspaper or website, emails within our inbox, songs within a playlist, or items on a restaurant menu. This type of verbal search differs from stereotyped environment of traditional text reading, in which your eyes scan from left to right along each line of text, from the top to the bottom of a page, as

sentences and paragraphs are scanned in sequential order (Rayner, 1998; Rayner, 2009). Instead, words in our modern environment are often arranged in complicated, variable, and widely spaced spatial configurations, as occurs on websites, magazines, maps, menus, calendars, schedules, and advertisements. Far less is known about how we locate word targets in these novel spatial arrangements.

Few studies have examined verbal and nonverbal search under fully-comparable experimental conditions. Paivio and Begg (1974) provide an early example of one such study. Participants were shown a cue followed by an array of 25 items (including a target and 24 distractors). Stimuli appeared as either written words or pictures, and the modality of both cue and array were manipulated in a factorial design. Behavioral results differed between the two within-platform task conditions, in which picture cues were followed by picture arrays versus when word cues were followed by word arrays. Participants were faster and more accurate on the nonverbal (picture) version of the task, compared to the verbal (word) version. This finding is consistent with a more general phenomena known as the "picture superiority effect," which describes how pictures are processed more rapidly than words, and subsequent memory for pictures is stronger as well (Nelson et al., 1976; Stenberg et al., 1995).

Paivio and Begg (1974) hypothesized that more efficient nonverbal search could reflect participants' engagement in a "parallel sweep" for pictures, versus a serial (item-by-item) search for words. They reasoned that visual features present in line drawing pictures would be discriminable in peripheral vision, allowing multiple objects to be screened simultaneously *via* covert mechanisms. In contrast, they noted that all words share similar sets of constituent letters, and may thus need to be foveated in order to be differentiated.

The primary goal of the current study is to leverage eye tracking to more directly evaluate Paivio and Begg's hypothesis. They inferred that parallel and serial search strategies (for objects and words, respectively) were being employed based on reaction time of pointing responses, but such strategies should also result in distinct patterns of eye movements. The use of serial search should be most obvious, as evidenced by a series of low-amplitude saccades to adjacent items in an array, until the target item is finally foveated. This yields the further prediction that serial search will result in roughly half of the items in the array being fixated before finding the randomly placed target.

In general it is only possible to foveate one item at a time, so any sort of parallel processing of multiple items must depend on the use of parafoveal (peripheral) vision. This should allow participants to "skip over" (i.e. pass by rather than fixate on) items that have been peripherally screened as not being the target. In eye tracking terms, this skipping behavior should be indexed by high amplitude saccades, and by fewer than half of array items being fixated on average (Seckin et al., 2016b).

In the current study we showed participants arrays of words, and on other trials arrays of object pictures, under reasonably comparable experimental conditions, while eye movements were monitored. Participants were provided with a cue (e.g. rabbit), and then tasked with locating a thematically associated target (e.g.

carrot) within each array. We expected to replicate the well-established picture superiority effect, such that manual touchscreen responses would be more rapid on nonverbal trials, compared to verbal trials. In keeping with Paivio and Begg's original hypothesis, we predicted that verbal trials would be more closely associated with serial viewing behavior (lower amplitude saccades and more items viewed), while nonverbal trials would be associated with parallel search behavior (higher amplitude saccades and fewer items viewed).

Additional Factors Governing Verbal and Nonverbal Search

In circumstances where participants engage in parallel search, this brings up a related question: by which basis would you skip over some objects but be drawn to foveate others? Would this be based on surface-level features, or deeper conceptual similarity to the target? Results from the commonly employed "visual world paradigm" (VWP) (Cooper, 1974), in which participants are presented with four object pictures and a spoken word cue that matches one of them, suggest both types of attributes can engender distraction. Object distractors of a similar shape to the target are viewed disproportionately longer in the VWP (Dahan and Tanenhaus, 2005; Huettig and Altmann, 2007; Yee et al., 2011), as are objects from the same taxonomic category (Sorensen and Bailey, 2007; Kalenine et al., 2012; Mirman and Graziano, 2012a; Mirman and Graziano, 2012b).

We included both types of distractors in the current paradigm, with additional controls not present in the standard VWP. Firstly, we independently manipulated shape and taxonomy, by including distractors from the same category as the target but with a dissimilar overall shape, and conversely distractors sharing shape but not category in common. This is necessary if both factors are to be examined separately, as members of the same taxonomic category almost always share visual features in common (Mirman et al., 2017), and often have the same overall shape (e.g. many tools have an elongation ending in a handle; many mammals have legs, arms, and a head). Secondly, we employed a full factorial design so that the cues and target arrays could be either words or pictures (unlike the original VWP, which is a strictly cross-platform task with word cues and picture arrays), allowing us to better disentangle the mechanisms involved in verbal versus nonverbal search. Based on previous results from the VWP, we predicted that both shape and taxonomic competitors would be viewed disproportionately on non-verbal trials. If so, this would suggest that nonverbal search is guided by a mixture of perceptual and conceptual processing of picture candidates (possibly in parallel).

In contrast, if Paivio and Begg's hypothesis is correct, and the search for well-spaced words is strictly serial, then neither shape nor taxonomy should affect processing for words in the parafoveal field. Rather, participants should be equally likely to view any sort of distractor, as they work their way around the array serially. It remains possible, however, that these factors would cause gaze to linger on words after foveation. We examined

these possibilities by evaluating both the number of distractors viewed and the duration of gaze upon viewing (Seckin et al., 2016b).

Finally, our full factorial design included not only within-platform conditions (i.e. verbal and nonverbal trials), but also cross-platform trials where word cues are followed by picture arrays and vice versa. These not only reflect every day circumstances in which words and pictures are intermixed (e.g. advertisements, signs, etc.), but allow us to compare our results to commonly employed cross-platform tasks such as the VWP. In general we expected the cross-platform results to be largely similar to their within-platform counterparts, such that participants would engage in a parallel search through pictures even when preceded by word cues, and to engage in serial search through words preceded by picture cues.

METHODS

Participants

Twenty-two undergraduates completed the study for course credit, including 13 males and 9 females, as approved by the Institutional Review Board at Cleveland State University. Participants reviewed and signed a written consent form prior to engaging in the study, and the study procedures were in accordance with the ethical standards of the Declaration of Helsinki. All participants were right-handed native English speakers, with an average age of 19.7 ± 2.6 years ($M ± SD$). All were screened for any history of medical or developmental conditions that could affect language or other domains of cognition.

Apparatus

A 20.5 × 11.5″ touchscreen monitor was used to present visual stimuli (1920 × 1080 resolution) and to collect touch responses. The screen was placed at an optimal distance for touch responses ($M ± SD = 20.4 ± 1.2″$), based on each participant's arm length. Eye movements were monitored and recorded using an EyeLink 1000 Plus eye tracking system (SR Research, Mississauga, ON, Canada). The eye tracking camera was placed above the participant's head, housed in a tower mount, so touch responses did not occlude the camera's view of the participant's eyes. A chin rest also housed in the tower mount minimized head movements. A nine-point calibration procedure was administered before the experiment to ensure accurate estimates of gaze position.

Materials

One hundred and seventy-five concrete object stimuli were selected for the eye tracking test (**Supplementary Table S1**). All items were presented as both pictures and written words throughout the course of the experiment. Picture stimuli were shaded gray scale drawings, drawn from the Rossion and Pourtois (2004) image set ($N = 44$) and internet image searches ($N = 131$). The pictures were scaled to 150 × 150 pixels (visual angle 4.5˚ × 4.5˚ from the average viewing distance of 20.4″). Word stimuli were the names for each item, written in lowercase 30-point Arial
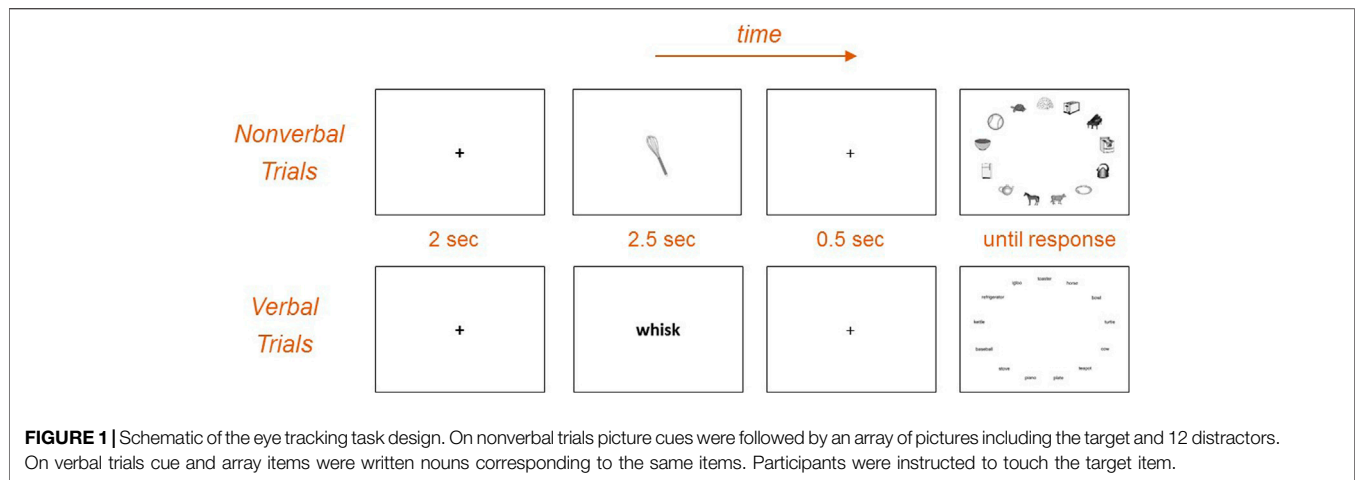
font. Words ranged in size from 41–226 pixels horizontally (1.2˚–6.8˚) and 22–37 pixels vertically (0.6˚–1.1˚). All stimuli were presented on a white background.

Twenty-four items were employed as cues, and paired with another 24 thematically-associated target items (e.g. "rabbit" and "carrot").[1] On each trial the target was accompanied by 12 distractors, whose taxonomic and visual relation to the target was systematically manipulated (**Supplementary Table S1**). Some distractors came from the same taxonomic category as the target, while others came from a different category. Specific categories employed included animals, body parts, foods, clothing, and tools. Some distractors had a similar overall shape to the target, for example by having outer boundaries that were long and thin ("asparagus," "flute"), round ("egg," "balloon"), or having an extended process ("key," "guitar"), while others had a different shape. Distractors were equally divided (3 each) into those sharing both category and shape in common with the target (e.g. "asparagus" for the target "carrot"), same-category but different-shape (e.g. "onion"), different-category but same-shape (e.g. "flute"), and different-category/different-shape (e.g. "shirt"). This allowed us to separately evaluate the influence of taxonomic and visual (shape) similarity.

Target and distractor words were balanced for psycholinguistic properties. Helmert repeated measures ANOVA contrasts showed that the target word did not differ from the 12 distractors each was paired with (**Supplementary Figure S1**) in word length (i.e. number of letters; $F_{(1,23)} = 0.33$, $p = 0.57$), lexical frequency (Lund and Burgess, 1996) ($F_{(1,23)} = 0.95$, $p = 0.34$), or orthographic neighborhood density (Balota et al., 2007) ($F_{(1,23)} = 0.19$, $p = 0.67$). The main effect of distractor type (same/different-category, same/different-shape) was also non-significant for word length ($F_{(3,69)} = 0.83$, $p = 0.48$), lexical frequency ($F_{(3,69)} = 1.23$, $p = 0.31$), and orthographic neighborhood density ($F_{(3,69)} = 2.09$, $p = 0.11$), suggesting the four types of distractor words were well balanced for psycholinguistic properties. Target and distractor pictures were balanced in visual complexity ($F_{(1,23)} = 0.08$, $p = 0.78$), as measured by compressed (JPEG) file size (Donderi and McFadden, 2005). Complexity did not significantly differ between the four types of picture distractors ($F_{(3,69)} = 1.29$, $p = 0.28$).

Cues were presented in the center of the screen, while targets and distractors were presented in an elliptical array with a horizontal axis of 1152 pixels (33.7˚) and a vertical axis of 878 pixels (26˚). This aspect ratio equates parafoveal acuity across positions when centrally-fixating (Iordanescu et al., 2011). At this degree of eccentricity from center, participants are typically able to detect whether an object is present (Thorpe et al., 2001) but are unable to discern their shapes (Nelson and Loftus, 1980). As such,

---

[1]As defined by Estes et al. (2011). "Thematic thinking: The apprehension and consequences of thematic relations," in *The psychology of learning and motivation: Advances in research and theory*, Vol. 54., editors. B. H. Ross and B. H. Ross. (San Diego, CA: Elsevier Academic Press), 249–294., a thematic relationship is "a temporal, spatial, causal, or functional relation between things that perform complementary roles in the same scenario or event."

FIGURE 1 | Schematic of the eye tracking task design. On nonverbal trials picture cues were followed by an array of pictures including the target and 12 distractors. On verbal trials cue and array items were written nouns corresponding to the same items. Participants were instructed to touch the target item.

the target could not be covertly identified from an initial central fixation, soliciting overt eye movements. Targets on each trial were equally distributed across the 13 possible array positions, as were the four types of distractors. Items in the array were spaced an average of 7.3° apart from one another.

Relative salience of each item in the array was examined, in order to ensure that low-level visual characteristics were not directing gaze preferentially towards the target or towards a particular class of distractors. Screen captured images of each word array and each picture array were converted into salience images using the Saliency Toolbox (Walther and Koch, 2006), based on the relative spatial distribution of intensity and line orientations across the screen. Average saliency values were averaged around the 150 pixels encompassing each of the 13 stimuli in the arrays. Helmert repeated measures ANOVA contrasts showed that the target items did not differ in salience from the distractors on each trial ($F_{(1,95)} = 0.028$, $p = 0.87$). The main effect of distractor type was also non-significant ($F_{(3,285)} = 1.4$, $p = 0.24$), suggesting the four types of distractors were comparable in salience.

Procedure

After a brief 9-point calibration procedure, participants completed the thematic eye tracking task. There were 96 trials in the eye tracking task, which took approximately 15 min to complete. On each trial the cue appeared in the center of the screen for 2.5 s, followed by a fixation cross for 0.5 s, followed by an elliptical array of 13 items, which remained on the screen until a touch response was detected (**Figure 1**). The array included the thematically-associated target and 12 distractors (including three each of same-category/same-shape, same-category/different-shape, different-category/same-shape, and different-category/different-shape).
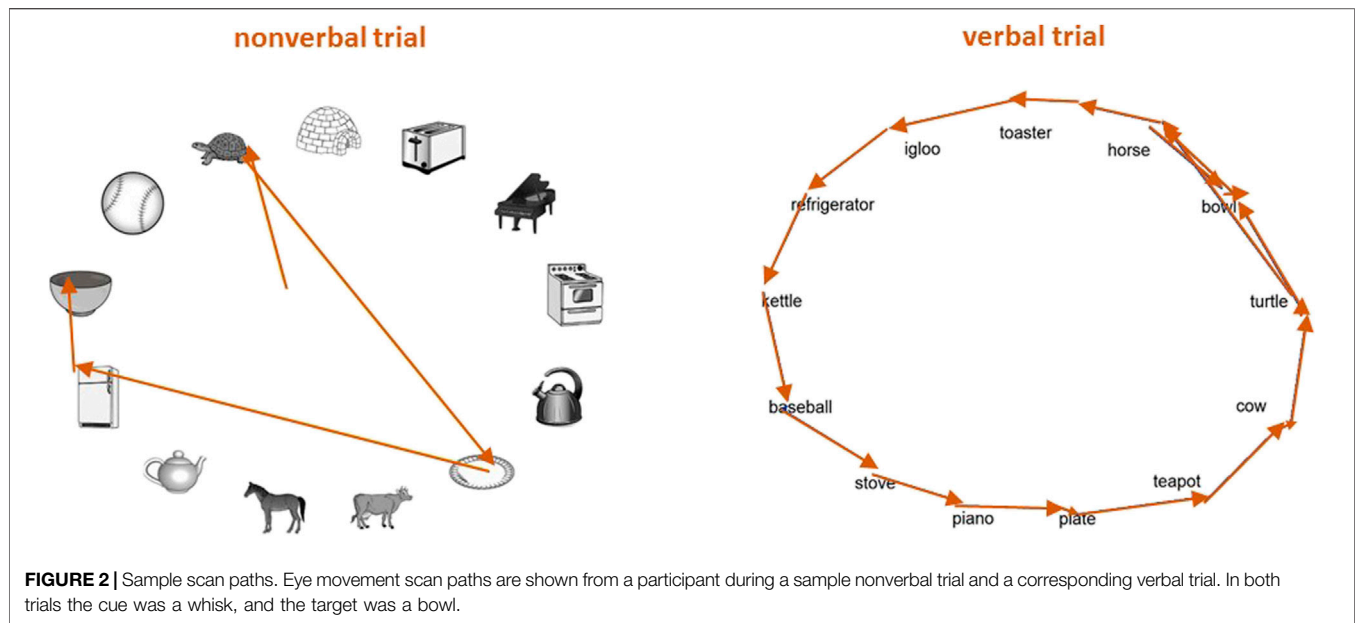
Stimulus modality varied across trials. On nonverbal trials ($N = 24$) cue and array were both pictures, and on verbal trials ($N = 24$) cue and array were both words (**Figure 1**). In the remaining cross-platform trials word cues were followed by picture arrays or vice versa ($N = 24$ each). The experiment was administered in blocks of six trials, each starting off with

an instruction screen indicating the relevant modality for that block (e.g. "Now you will match words with pictures") along with an example cue and array. Stimulus modality was changed after each block, preventing results in any modal version from being confounded by practice, fatigue, or other ordering effects. The first block of six trials were nonverbal (picture-to-picture), followed by a block of cross-platform word-to-picture trials, followed by a block of verbal trials (word-to-word), followed by a block of cross-platform picture-to-word trials. There were four cycles of this repeated block pattern throughout the experiment, resulting in 96 total trials. Each cue and target appeared four times throughout the experiment: twice as word stimuli (once within-platform, once cross-platform), and twice as picture stimuli (once within-platform, once cross-platform). Repetition of stimuli was balanced across blocks and cycles, such that the stimuli in each modal block were matched for the number of prior exposures (avoiding potential confounds between modality and degree of exposure). All participants completed these trials, blocks, and cycles in the same pseudo-randomized order.

Eye Movement Acquisition and Coding

Eye movements were recorded at a sampling rate of 500 Hz, in epochs starting with the onset of the array and ending with each touch response. The EyeLink Data Viewer software package was used to classify each continuous recording into blinks, saccades, and fixation. Periods without a detectable pupil were classified as blinks. Saccades were identified as eye movements that crossed motion (0.15°), velocity (30°/s), and acceleration thresholds (8,000°/s²). Any remaining period lasting ≥40 ms was classified as a fixation.

Areas of interest (AOIs) were constructed around each of the 13 items in the array (**Supplementary Figure S1**), and were used to classify fixations and saccades. An imaginary outer iso-acuity ellipse (with the same proportions as the stimulus array) was plotted, which extended to the left, right, top, and bottom outer borders of the screen. An inner concentric ellipse was also plotted, at the same distance from the stimulus ellipse but located closer to the center of the screen. Straight lines were then plotted between

FIGURE 2 | Sample scan paths. Eye movement scan paths are shown from a participant during a sample nonverbal trial and a corresponding verbal trial. In both trials the cue was a whisk, and the target was a bowl.

the inner and outer boundary ellipses, such that each line bisected the space between adjacent items in the array. The inner and outer points of each bisecting line were then connected to create a series of trapezoidal AOIs, each centered on an item in the array. This maximized our ability to assign mildly mis-calibrated fixations to the nearest item in the array. Fixations falling outside the AOIs (in the center or peripheral corners of the screen), and saccades with a starting or ending point outside the AOIs, were excluded from analysis. Eye movements on inaccurate trials were also excluded from analysis (<6% of trials per condition).

Percent viewing time for each item was calculated as the summed duration of all fixations in that item's AOI, on that trial, divided by the total length of time spent in all AOIs on that trial. Percent viewing times were then used as a dependent variable in statistical analysis.

Statistical Analysis

Normality of the dependent variables, including touch response accuracy and reaction times as well as percent dwell times, was assessed via Kolmogorov-Smirnov tests. All variables were found to be normally distributed [$p > 0.05$ in all task conditions (verbal, non-verbal, and cross-platform trials)], with the exception of accuracy ($p < 0.05$ in all task conditions). Accuracy data were therefore examined via non-parametric Wilcoxon signed-ranks tests. Reaction times were analyzed via paired-samples t-tests, and percent viewing times on distractors were analyzed with 2 × 2 repeated-measures ANOVAs across shape and taxonomic category. A series of additional factors for percent viewing times (e.g. lexical frequency) were analyzed at the trial level using generalized linear mixed-effects models (GLMMs). More details are presented throughout the *Results*. All statistical analyses were conducted with SPSS v25 (IBM Corp, Armonk, NY, United States) and SAS version 9.2 (SAS Institute Inc., Cary, NC, United States).

RESULTS

Results from nonverbal and verbal trials will first be described in detail, followed by a summary of cross-platform results.
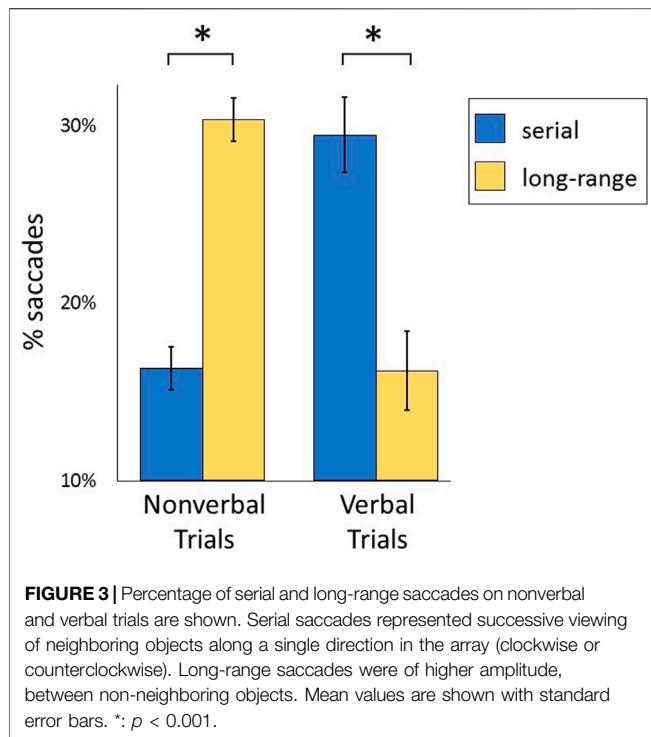
Touch Responses

Accuracy was higher on nonverbal trials ($M \pm SD = 97.54 \pm 3.56\%$) compared to verbal trials ($94.32 \pm 5.07\%$), according to Wilcoxon signed-rank test ($Z = 2.29$, $p = 0.022$, $\eta^2 = 0.25$). Likewise, reaction times were faster on nonverbal trials ($2,805 \pm 638$ ms) compared to verbal trials ($3,138 \pm 585$; $t_{(21)} = 3.49$, $p = 0.002$, $\eta^2 = 0.367$). Reaction times on inaccurate trials were excluded from this analysis.

Saccade Classification

Scan paths such as the ones shown in **Figure 2** demonstrate that participants tended to engage in a mixture of low and high amplitude saccades on nonverbal trials, but tended to adopt a more stereotyped and exhaustive "serial search" strategy (Seckin et al., 2016a; Seckin et al., 2016b) on verbal trials. In order to quantify these patterns, saccades were classified into three types based on their start and end points: within-item saccades starting and ending in the same AOI, local saccades where the start and end points are in adjacent AOIs (corresponding to neighboring items along the array), and long-range saccades where the end point is in a non-adjacent AOI. Local saccades that were preceded by another local saccade in the same direction (clockwise or counterclockwise) were further coded as being serial saccades, with the remainder of local saccades being labeled as non-serial. The mean amplitude of within-item saccades was $1.31 \pm 0.17°$, local saccades (including serial) had a mean amplitude of $4.38 \pm 0.19°$, and long-range saccades had an average amplitude of $11.34 \pm 0.67°$.

As shown in **Figure 3**, serial saccades were more prominent on verbal trials ($29.45 \pm 9.96\%$ of all saccades) than nonverbal trials

FIGURE 3 | Percentage of serial and long-range saccades on nonverbal and verbal trials are shown. Serial saccades represented successive viewing of neighboring objects along a single direction in the array (clockwise or counterclockwise). Long-range saccades were of higher amplitude, between non-neighboring objects. Mean values are shown with standard error bars. *: $p < 0.001$.

FIGURE 4 | Each trial included distractors that were from the same or different category as the target, and had the same general shape or a different shape from the target. Average time spent fixating each class of distractor is shown as a percentage of total fixation time during that trial. *: $p < 0.001$.

$(16.31 \pm 5.66\%; t_{(21)} = 7.21, p < 0.001, \eta^2 = 0.712)$. Conversely, long-range saccades were more prominent on nonverbal trials $(30.31 \pm 5.66\%)$ compared to verbal trials $(16.34 \pm 7.37\%; t_{(21)} = 11.75, p < 0.001, \eta^2 = 0.868)$. The percentage of within-object saccades did not differ according to modality $(t_{(21)} = 0.68, p = 0.507, \eta^2 = 0.021)$, nor did the percentage of non-serial local saccades $(t_{(21)} = 0.11, p = 0.910, \eta^2 = 0.001)$.

Fixations on Each Distractor Type

Participants viewed fewer items on nonverbal $(5.55 \pm 0.99$ out of 13 items) than verbal trials $(6.47 \pm 1.27; t_{(21)} = 5.27, p < 0.001, \eta^2 = 0.570)$. In addition, each item viewed was also fixated for a shorter duration on nonverbal trials $(178 \pm 66$ ms) compared to verbal trials $(212 \pm 55$ ms; $t_{(21)} = 2.92; p = 0.008; \eta^2 = 0.289)$. Percent viewing time for each item was calculated as the summed duration of all fixations in that item's AOI, on that trial, divided by the total length of time spent in all AOIs on that trial. Percent viewing time on each class of distractor is shown in **Figure 4**. An omnibus repeated measures ANOVA including the fixation data from both verbal and nonverbal trials revealed significant interactions between modality and category $(F_{(1,21)} = 32.66, p < 0.001, \eta^2 = 0.356)$, and between modality and shape $(F_{(1,21)} = 17.52, p < 0.001, \eta^2 = 0.256)$. The magnitude of categorical and shape-based capture of gaze thus depends upon the modality of stimulus. When viewing time on distractors was analyzed solely within nonverbal trials, the main effect of category $(F_{(1,21)} = 46.14, p < 0.001, \eta^2 = 0.558)$ and shape $(F_{(1,21)} = 36.22, p < 0.001, \eta^2 = 0.476)$ were both significant. In contrast, neither category $(F_{(1,21)} = 0.21, p = 0.665, \eta^2 = 0.002)$ nor shape $(F_{(1,21)} = 0.03, p = 0.877, \eta^2 < 0.001)$ induced capture on verbal trials.
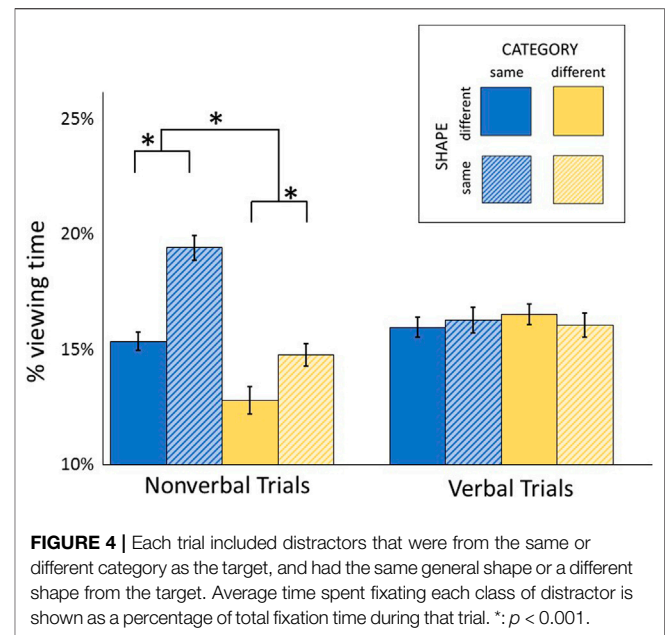
The prevalence of long-range over serial saccades suggests participants employed covert attentional mechanisms during nonverbal search, which may have contributed to capture on those trials. A parafoveal mechanism of capture would draw gaze towards an object distractor in the periphery, making it more likely to be viewed on each trial (Seckin et al., 2016b). To assess this possibility, we calculated the number of distractors viewed on each trial, such that at least one fixation fell within their respective AOIs.

Modeling Fixation Predictors

Neither shape nor category affected percent viewing times on words, but participants were slower and less accurate on verbal trials, suggesting that other perceptual or conceptual factors may have governed word-based interference. Univariate and multivariable generalized linear mixed-effects models (GLMM) were constructed to examine which distractor attributes were associated with greater percent viewing times (as the dependent outcome measure). Predictors included distance from target (number of intervening items between the distractor and target item along the array), saliency, category (same/different), shape (same/different), word length (number of letters in each distractor word), word length difference from target (computed as the absolute value of the length of distractor word minus the length of the target word), and lexical frequency. Trials were nested within participants, with random-effects accounting for clustering. Random intercept was included with unstructured covariance structure. Distractor percent viewing times were included as outcome measures in separate GLMMs for verbal and nonverbal trials. Bonferroni correction was applied to the significance threshold in univariate models, to control for elevated type 1 error when testing seven predictors independently $(\alpha = 0.007)$.

Rather than shape and category, lexical predictors accounted for variance on verbal trials (**Table 1**). Participants spent increased

TABLE 1 | Predictors of viewing times on verbal trials. Parameter estimates (and the standard error associated with each) are shown when considered in isolation (univariate) and in combination with the other predictors (multivariable).

	Univariate		Multivariable	
	Estimate (SE)	p-value	Estimate (SE)	p-value
Distance to target	−0.14 (0.03)	<0.001*	−0.14 (0.03)	<0.001[‡]
Salience	1.03 (1.09)	0.344	1.36 (1.09)	0.214
Shape	0.07 (0.20)	0.707	0.08 (0.20)	0.673
Category	−0.22 (0.20)	0.266	−0.38 (0.20)	0.056
Absolute word length	0.46 (0.05)	<0.001*	0.54 (0.06)	<0.001[‡]
Word length difference from target	−0.16 (0.06)	0.005*	−0.36 (0.06)	<0.001[‡]
Lexical frequency	−0.24 (0.05)	<0.001*	−0.07 (0.06)	0.242

*: $p < 0.007$ (Bonferroni correction). [‡]: $p < 0.05$.

TABLE 2 | Predictors of percent viewing times on nonverbal trials.

	Univariate		Multivariable	
	Estimate (SE)	p-value	Estimate (SE)	p-value
Distance to target	−0.12 (0.03)	<0.001*	−0.15 (0.03)	<0.001[‡]
Salience	−0.54 (0.29)	0.059	−0.37 (0.29)	0.202
Shape	0.99 (0.18)	<0.001*	1.11 (0.18)	<0.001[‡]
Category	1.06 (0.18)	<0.001*	1.13 (0.18)	<0.001[‡]
Absolute word length	−0.004 (0.05)	0.930	0.03 (0.05)	0.521
Word length difference from target	−0.06 (0.05)	0.280	−0.08 (0.06)	0.144
Lexical frequency	0.03 (0.05)	0.521	0.09 (0.05)	0.073

*: $p < 0.007$ (Bonferroni correction). [‡]: $p < 0.05$.

time viewing words with more letters (absolute word length), words of a similar length to the target, and lower frequency words. Longer words also tend to be less frequent, causing the latter to be non-significant in the multivariable model. Lexical predictors were not associated with viewing times on nonverbal trials (**Table 2**), where shape and category again proved to be the relevant factors. Participants spent more time viewing items that were closer to the target (distance to target) on both trial types. Salience did not affect viewing times, and did not interact with other factors in the multivariable models.

Results From Cross-Platform Trials

There were two types of cross-platform trials: word cues followed by picture arrays (word-to-picture), and picture cues followed by word arrays (picture-to-word). Accuracy was 95.26 ± 6.07% on word-to-picture trials and 96.96 ± 4.66% on picture-to-word trials. Reaction times were 2,780 ± 613 ms on word-to-picture trials, and 2,910 ± 535 ms on picture-to-word trials. When compared directly to their within-platform counterparts, word-to-picture accuracy ($Z = 1.5$; $p = 0.135$; $\eta^2 = 0.107$) and reaction times ($t_{(21)} = 0.29$; $p = 0.772$; $\eta^2 = 0.004$) did not differ significantly from those on nonverbal (i.e. picture-to-picture) trials, but responses on picture-to-word trials were significantly more accurate ($Z = 2.3$; $p = 0.021$; $\eta^2 = 0.252$) and faster ($t_{(21)} = 2.47$; $p = 0.022$; $\eta^2 = 0.225$) compared to performance on verbal trials.

Long-range saccades were more prevalent (29.37 ± 5.62%) than serial saccades (17.53 ± 5.85%; $t_{(21)} = 5.29$, $p < 0.001$, η^2 0.571) on word-to-picture trials (**Supplementary Figure S2**). These patterns were similar to those present on nonverbal
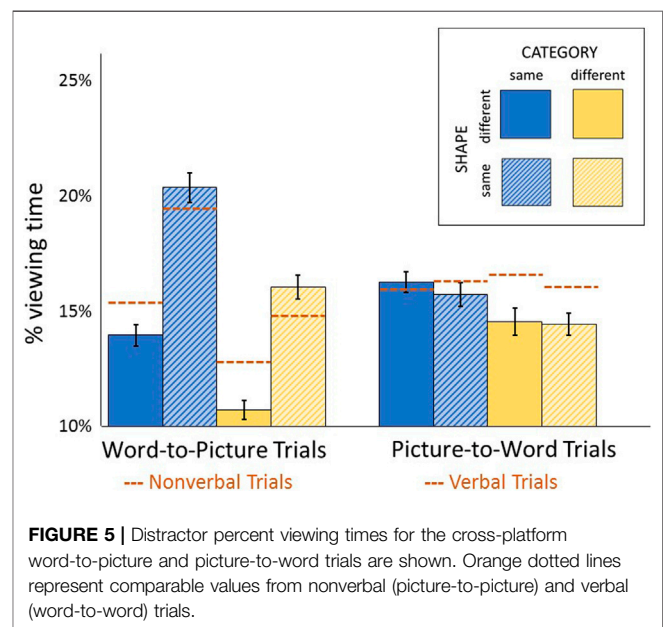

FIGURE 5 | Distractor percent viewing times for the cross-platform word-to-picture and picture-to-word trials are shown. Orange dotted lines represent comparable values from nonverbal (picture-to-picture) and verbal (word-to-word) trials.

trials (**Figure 3**), as evidenced by a non-significant interaction between trial type (word-to-picture vs nonverbal) and saccade type (serial vs long-range) ($F_{(1,21)} = 1.02$, $p = 0.323$, $\eta^2 = 0.039$). Conversely, participants generated more serial (29.1 ± 8.07%) than long-range saccades (17.15 ± 6.03%; $t_{(21)} = 4.32$, $p < 0.001$, $\eta^2 = 0.471$) during picture-to-word trials, closely

mirroring saccadic behaviors on verbal trials ($F_{(1,21)}$) = 0.29, p = 0.595, 0.008).

Distractor viewing times were then examined on cross-platform trials (**Figure 5**). Participants spent a similar percentage of time viewing distractors on word-to-picture (61.68 ± 6.48%) and nonverbal trials (62.31 ± 4.34%; $t_{(21)}$ = 0.50; p = 0.624; η^2 = 0.012), but spent proportionately less time viewing distractors on picture-to-word trials (60.96 ± 7.39%) compared to verbal trials (64.79 ± 5.97%; $t_{(21)}$ = 3.27; p = 0.004; η^2 = 0.337). Participants viewed each item on picture-to-word trials for an average duration of 174 ± 53 ms, compared to 213 ± 55 ms on verbal trials ($t_{(21)}$ = 4.41; p < 0.001; η^2 = 0.481), suggesting that word distractors were rejected more efficiently when preceded by a picture cue.

Similar to nonverbal trials (**Figure 4**), there was a main effect of category ($F_{(1,21)}$ = 57.46, p < 0.001, η^2 = 0.458) and shape ($F_{(1,21)}$ = 197.41, p < 0.001, η^2 = 0.728) on word-to-picture trials (**Figure 5**). When directly compared via interaction terms, the shape effect was of higher amplitude on word-to-picture compared to nonverbal trials ($F_{(1,21)}$ = 39.25, p < 0.001, η^2 = 0.275), while the category effect was of equivalent amplitude in both trial types ($F_{(1,21)}$ = 0.04, p = 0.840, η^2 = 0.001). Although the shape effect was non-significant on picture-to-word trials ($F_{(1,21)}$ = 0.91, p = 0.350, η^2 = 0.003), the category effect was significant ($F_{(1,21)}$ = 13.76, p = 0.001, η^2 = 0.119) and of higher amplitude compared to verbal trials ($F_{(1,21)}$ = 12.84, p = 0.002, η^2 = 0.064). The number of distractors viewed was calculated to assess parafoveal contributions to the category effect. Participants viewed a greater number of same-category distractors (compared to different-category distractors) on word-to-picture trials ($F_{(1,21)}$ = 30.42, p < 0.001, η^2 = 0.592), but not on picture-to-word trials ($F_{(1,21)}$ = 0.64, p = 0.433, η^2 = 0.029).

GLMMs were conducted to identify additional factors related to distractor viewing times on cross-platform trials. Shape and category were significant predictors on word-to-picture viewing times, as was the case for nonverbal trials (**Supplementary Table S2**). Unlike nonverbal trials, absolute word length was also significantly associated with word-to-picture viewing times in the univariate model. This effect did not achieve significance in the multivariable model (p = 0.055) due to shared variance with other predictors (shape, and word length difference from target). In contrast, absolute word length and lexical frequency were associated with viewing times on picture-to-word trials (**Supplementary Table S3**), as was the case for verbal trials.

DISCUSSION

Participants in the current study engaged in rigorous verbal and nonverbal search tasks under comparable conditions, while eye movements were monitored. Participants were highly accurate in all task conditions, suggesting that the verbal and nonverbal tasks were matched in difficulty (Thierry and Price, 2006). There were, however, subtle but significant differences in performance between platforms. Nonverbal search was 11% faster and 3% more accurate than verbal search, consistent with the well-established picture superiority effect (Nelson et al., 1976).

Eye movements revealed some of the mechanisms underlying picture superiority in the context of visual search. Verbal search was dominated by serial eye movements, as evidenced by a series of clockwise or counter-clockwise low-amplitude saccades to adjacent words in the array. In contrast, the more efficient nonverbal search was associated with more frequent long-range saccades to non-adjacent items, and fewer total items being viewed on each trial.

These discrepant strategies are supportive of Paivio and Begg's initial hypothesis that, unlike words which require serial search, we are able to engage in a "parallel sweep" for objects (Paivio and Begg, 1974). Only one object can be foveated at a time, so parallel viewing depends upon the effective use of peripheral vision, leveraging covert attention to scan small clusters of objects simultaneously (Smith et al., 2014; Fluharty et al., 2016). Saccadic patterns observed in the nonverbal task are consistent with this interpretation. After executing a long-range saccade to a sector of the array, participants seem able to simultaneously screen not only the foveated item (overtly) but also adjacent pictures in the array (covertly). Participants can then execute a local saccade to the target if it happens to be present within the effective parafoveal field. If the target is not present, the participant can skip over covertly screened adjacent pictures, and instead execute another long-range saccade to a new sector of the array.

As further evidence that object features were discernable in the periphery, participants were more likely to fixate on distractors from the same category and of a similar shape to the target on nonverbal trials (i.e. viewed a greater number of those distractors on each trial). Participants fixated on fewer total distractor pictures than distractor words, however, and identified picture targets more rapidly than word targets, suggesting that employment of covert attention was largely advantageous to nonverbal search. In addition to parallel capacity, covert attention can be directed faster than overt eye movements can be executed (Posner et al., 1980), which may further serve to accelerate nonverbal search and promote picture superiority in search contexts.

Although word stimuli in the current study were presented at roughly the same size as picture stimuli, and placed at the same eccentricity along each array (7.3°), results from verbal trials suggest that word features were not processed to any meaningful extent in peripheral vision. Participants viewed an average of 6.5 out of the 13 words on each trial before initiating a touch response, which is precisely what would be predicted in a scenario where the target is placed randomly and every item needs to be directly foveated in order to be rejected. The finding that peripheral pictures but not words were discernable in our arrays is consistent with the general principle that the size of the effective parafoveal field (also known as "perceptual span") depends on the modality of the stimuli being viewed. Estimates of the effective field size for objects in scenes range from 4–10° (Parker, 1978; Henderson et al., 2003), while the effective field for letters and words is limited to 3–4 character spaces to the left of fixation and 14–15 spaces to the right of fixation, at least in the context of passage reading (Rayner, 2009).

A variety of visual attributes inherent to written English words likely contribute to their being difficult to discern in the periphery. All words have a similar overall shape, with longer horizontal than vertical axes. Whereas the visuospatial arrangement of object features is meaningful (e.g. the "business end" of objects is often distal, and many are symmetrical), letters are instead ordered via a

symbolic mapping with temporal speech sounds. Theoretically, each letter within a word may be any of 26 in the English alphabet, but in practice the six most frequent letters represent over half of those encountered in text (Zim, 1966; Soanes and Stevenson, 2004; Norvig, 2012), resulting in substantial letter overlap across words. Letters themselves are comprised of a limited set of features, including curves, lines, and dots, and are closely spaced within each word. Physical similarity between letters acts to further reduce the effective span (Rayner and Fisher, 1987). Put more simply, all words look the same, and the lack of distinguishing features renders them less actionable in the periphery.

When visual features in array items are highly overlapping and "inseparable," a serial rather than parallel strategy becomes more effective (Treisman and Gelade, 1980), as was apparently the case for words in the current study. Serial viewing is also dominant in traditional reading contexts, during which some words are skipped over, but the majority are sequentially fixated in passages (Rayner, 1998; 2009). We should point out that the use of a serial strategy is not mutually exclusive with use of covert attention (Cave and Wolfe, 1990). Typical adults covertly screen adjacent words during passage reading, particularly to the right of those being fixated, which conveys a "preview benefit" (Briihl and Inhoff, 1995; Miellet and Sparrow, 2004). Words in the current study were spaced further apart than in passages, however, and were vertically displaced from one another, which drastically curtails peripheral processing of word and letter content in horizontal lines and strings (Inhoff and Briihl, 1991; Pollatsek et al., 1993).

The current findings may therefore be more relevant for contemporary everyday reading behaviors, where words are distributed in variable configurations rather than traditional text passages. The content of words is less likely to draw gaze in such configurations, as further evidenced by the finding that banner advertisements are no more likely to be viewed when they are conceptually-congruent with web sites (Hervet et al., 2011). Instead, important verbal material may be more effectively highlighted by low-level manipulations such as font, color, and placement (Lohse, 1997; Kuisma et al., 2010; Slattery and Rayner, 2010). The current results may also help explain why pictures are more likely to draw gaze to advertisements than text (Pieters and Wedel, 2004), as they are identified in a wider effective parafoveal field.

Different Factors Capture Gaze During Verbal and Nonverbal Search

Saccadic patterns indicated that objects can be screened in parallel, so additional fixation analyses were conducted to determine why some objects were skipped over while others were more effective at capturing attention. Distractors similar in shape to the target object, as well as distractors from the same taxonomic category, were viewed disproportionately during nonverbal search. These findings are consistent with previous results from the VWP (Dahan and Tanenhaus, 2005; Huettig and Altmann, 2005; Huettig and Altmann, 2007; Sorensen and Bailey, 2007; Yee et al., 2009; Yee et al., 2011; Kalenine et al., 2012; Mirman and Graziano, 2012a; Mirman and Graziano, 2012b), and demonstrate that both factors are influential in entirely nonverbal task conditions as well. Importantly, taxonomic distractors were effective at capturing

gaze even when they had a dissimilar shape to the target, suggesting that taxonomic competition between pictures was not limited to shared perceptual characteristics. This finding is consistent with theoretical accounts in which category coordinates are co-activated via conceptual priming (Collins and Loftus, 1975; Masson, 1995; Hutchison, 2003; Proklova et al., 2016). In summary, the relative efficiency of nonverbal search appears to be facilitated by parallel screening; objects which are perceptually or conceptually similar to the target are more likely to be foveated, while unrelated objects are more likely to be skipped over.

In contrast, the serial strategy taken by participants during verbal trials ensured that no type of distractor was more likely to be viewed. Instead, adjacent words were viewed one by one until the target was identified, regardless of their composition. It remained theoretically possible that shape and taxonomy could still capture gaze through foveal mechanisms, such that gaze would linger on those types of distractors after being fixated (Seckin et al., 2016b). These distractors were viewed for the same duration as unrelated words, however, refuting the idea that either shape or taxonomy could act as an attentional "sinkhole."

Lack of shape effects for words is consistent with results from Huettig and McQueen (2007), who also failed to find any influence of shape when they conducted an all-verbal adaptation of the VWP. As described in the picture superiority literature (Nelson et al., 1976; Nelson et al., 1977; Sperber et al., 1979), features including shape are directly represented in pictures, but would need to be retrieved from long-term memory when viewing words, potentially explaining why shape did not influence verbal search.

The absence of a taxonomic effect during verbal search is more perplexing, as taxonomic priming is consistently observed when processing pairs of words in psycholinguistic paradigms, resulting in faster reaction times, lower amplitude N400 electro- and magnetoencephalographic potentials, and hemodynamic suppression for words from the same category compared to unrelated words (Bentin et al., 1985; Rossell et al., 2003; Wheatley et al., 2005; Lewis et al., 2015). The most obvious explanation would be that lack of peripheral information for words may have therefore mitigated their ability to capture attention, but this explanation is unsatisfying as taxonomic capture was observed on cross-platform trials with the exact same word arrays (see next section for discussion of cross-platform results). We are therefore left with the curious finding that taxonomic capture emerges when either cues or targets are pictures, but evaporates in entirely verbal task circumstances.

Taxonomic processing was somehow marginalized during verbal search in the current study, suggesting that, rather than being obligatory, automatic, and ubiquitous, at least some types of lexico-conceptual information are evoked in a task-dependent manner. This is consistent with (Friedman and Bourne's, 1976) contention that "... multiple codes or representations do not exist to be activated by the appropriate stimuli, but rather the stimuli themselves embody levels of information which are encoded and used as needed." According to their perspective (which was motivated by a levels-of-processing framework), words and objects may each be flexibly processed at different levels of depth, extracting qualitatively different types of information according to the demands of the situation. Word category becomes pertinent when search is guided

by picture cues, but not when guided by word cues, suggesting that pictures somehow elicit a taxonomic "mindset" that is applied to the search task. This possibility requires further study.

The absence of taxonomic effects during verbal search begs the question as to what sorts of lexico-conceptual information, if any, were accessed as participants viewed words in the current study. Alternatively, were participants merely screening for the sequence of letters that comprise the target word, constraining processing to a superficial orthographic level rather than engaging in lexico-conceptual access? A purely orthographic strategy would be efficient, as participants could rapidly reject each distractor when the first letter does not match the target, but the results from GLMM analyses suggest this was not the case. Lexical frequency was found to significantly modulate fixation durations, such that low frequency word distractors were viewed for longer, which is a common finding in studies of passage reading (Rayner, 1998; Kliegl et al., 2004; White, 2008). This suggests that stored lexical representations corresponding to each word distractor were indeed accessed during verbal search.

Viewing time GLMMs revealed additional physical characteristics of word distractors which caused gaze to linger. The absolute length of words affected gaze, such that words with more characters were viewed for longer (irrespective of their relationship with the target word), which is also common when reading text passages (Kliegl et al., 2004). More interestingly, words closer in length to the target were also viewed for a greater duration. This suggests that participants were employing some sort of whole-word recognition strategy (Ellis and Young, 1988) as the general form of the target word was employed as a template during verbal search, providing top-down guidance (Wolfe et al., 2004; Schmidt and Zelinsky, 2009).

In summary, qualitatively different factors were found to govern verbal compared to nonverbal search, with the former being modulated by orthographic and lexical factors such as word length and frequency, while the latter was affected by the taxonomic category and shape of the picture distractors.

Characteristics of Cross-Platform Search

In addition to within-platform conditions, the factorial design of the current study also included blocks of cross-platform trials where word cues were followed by picture arrays and vice versa. As expected, saccadic patterns indicated that the strategies employed in cross-platform search were largely dictated by the modality of the stimulus arrays. Long-range saccades were prevalent in word-to-picture trials, as was the case for nonverbal trials. Likewise, picture-to-word search was dominated by serial viewing of word stimuli, as was the case on verbal trials. These strategies are plausibly driven by the quality of parafoveal information available in the display, allowing pictures but not words to be viewed in parallel. Serial search is inherently slower, as reflected by greater reaction times on picture-to-word as compared to word-to-picture trials.

There were, however, subtle differences in performance on cross-platform compared to within-platform trials. Given that the physical image of the array on the retina was identical on cross-platform trials and their within-platform counterparts, these differences are attributable to the platform of the cue being held in working memory. Although Paivio and Begg (1974) found no differences in search times for word targets preceded

by picture cues versus word cues, we found that search was 7% faster and 3% more accurate on cross-platform picture-to-word trials (compared to verbal trials). Analysis of fixation times revealed that participants spent proportionately less time viewing distractors on picture-to-word trials compared to verbal trials, further indicating that top-down guidance from picture cues was more effective when searching for words.

Previous studies employing picture targets found they were identified more rapidly when preceded by picture cues, compared to word cues (Paivio and Begg, 1974; Wolfe et al., 2004; Schmidt and Zelinsky, 2009). In the current study, there were no significant differences in reaction times or accuracy on word-to-picture compared to picture-to-picture (nonverbal) trials. This is unlikely to represent a type II error, as the effect was not only non-significant but the reaction time means were patterned in the opposite direction as predicted, being 25 ms faster on word-to-picture trials.

Picture cueing benefits in previous studies were interpreted as reflecting the creation of a more effective guidance template, which could then be applied during the search for a subsequent picture target. Schmidt and Zelinsky (2009) found that the more specific the cue was, the larger the benefit, with category names being the least effective cue, followed by specific object names, and picture cues being the most effective. Unlike previous studies, picture cues in the current study were thematic associates of the subsequent target, rather than identically-matching pictures. As such, the resultant guidance templates may have been less specific, failing to convey a substantial benefit for picture cues (as compared to word cues) when searching for picture targets.

In general, the same factors that captured gaze on within-platform trials tended to elicit distraction in their cross-platform counterparts: shape and category were influential on word-to-picture trials (as was the case for nonverbal trials), while word length and frequency were influential on picture-to-word trials (as was the case for verbal trials). There were some notable exceptions, however, demonstrating that the mixed cues on cross-platform trials affected subsequent distraction. Firstly, the shape effect was of higher amplitude on word-to-picture compared to entirely nonverbal trials. More strikingly, other factors were found to selectively affect viewing times only on cross-platform trials. Although performance on nonverbal trials was unaffected by any orthographic or lexical factors, word length significantly modulated gaze on word-to-picture trials. Participants viewed object pictures with longer names for a greater amount of time, suggesting that those names were covertly accessed to an extent that was consequential for performance, but only when those pictures were preceded by a word cue. Covert naming effects have also been observed when viewing pictures in the cross-platform VWP (Dahan et al., 2001).

Taxonomic capture was of equivalent amplitude on word-to-picture and nonverbal trials, and was far greater on picture-to-word compared to verbal trials (with the effect being non-significant in the latter). Although taxonomic capture was observed in both types of cross-platform trials, there appeared to be different underlying mechanisms. Participants spent a disproportionate percentage of time viewing same-category distractors on both types of cross-platform trials, but only viewed a greater number of same-category distractors on word-to-picture trials. This indicates

parafoveal involvement on word-to-picture trials, such that gaze was drawn towards same-category pictures in the periphery, as was the case on nonverbal trials. In contrast, lack of peripheral discriminability meant that participants were no more likely to view same-category words on picture-to-word trials, so preferential viewing of lexical competitors was more likely based on foveal mechanisms.

Two conclusions can be drawn based on the results from cross-platform trials. Firstly, although it would be reasonable to expect that within-platform cues would provide more direct and therefore stronger guidance, the inclusion of cross-platform cues does not appear to be deleterious for visual search. Word-to-picture search was equally efficient as picture-to-picture search, while picture-to-word search was slightly more efficient than word-to-word search. Secondly, whereas separate and distinct sets of factors were found to create distraction during verbal and nonverbal search, these sets overlapped during cross-platform search. Orthographic and lexical factors affected word viewing during verbal search, while taxonomy and shape affected picture viewing during nonverbal search, but a mixture of both sets of factors were influential in the cross-platform conditions. This makes sense given that both words and pictures are present, and must be considered in relation to one another, on each cross-platform trial. Word-to-picture search appears to have encouraged verbal mediation while viewing objects, as evidenced by a word length effect. Likewise, picture-to-word search apparently encouraged nonverbal mediation while viewing words, as evidenced by taxonomic capture (which was not present in the entirely verbal condition). This raises the possibility that many of the most commonly-administered language tasks such as the VWP and the picture-word interference task (Schriefers et al., 1990) may also encourage nonverbal mediation (*via* the inclusion of picture stimuli), resulting in effects which would otherwise not be present in a purely linguistic design.

Limitations

The current results indicate that parallel viewing is one of the core mechanisms underlying picture superiority in the context of visual search. Picture superiority is undoubtedly supported by additional mechanisms, however, as superiority effects are observed in situations where parallel processing is not possible, such as when judging solitary items (Sperber et al., 1979; Bajo, 1988; Stenberg et al., 1995). Future studies, particularly those involving novel paradigms and convergent techniques, may further our understanding of the differential mechanisms underlying word and object recognition.

We tried to place words and objects on an equal footing in the current eye tracking paradigm, equating them for size and position in each array. The effective parafoveal field is larger, however, for objects than for words, so there was greater opportunity for parafoveal involvement on nonverbal trials. Although one could argue that these differences in effective span are an important aspect of verbal and nonverbal search in real world circumstances, future studies could also attempt to equate both types of modal arrays for effective span, in order to address a slightly different set of research questions.

We also repeated the same set of cues, targets, and distractors across platforms. Again, this was done in the interest of placing words and objects on equal footing, as the exact same items were present on verbal, nonverbal, and cross-platform trials, which eliminated a host of potential confounds. Repetition surely affects processing however, and "long-term" repetition (i.e. with lag times that fall outside the span of working memory) has been found to have differential effects across stimulus modalities, with effects being strongest in cross-platform conditions (Stenberg et al., 1995). It would be interesting to observe whether the effects reported in this study would change if repetition was either removed altogether (Schmidt and Zelinsky, 2009), or systematically manipulated.

We manipulated the shape of pictures in the current study, in order to disentangle perceptual from conceptual influences on visual search, and to further isolate the mechanisms of taxonomic capture (by removing the visual similarity between category competitors). Our shape manipulation was based on the overall outline of objects, however, which were crudely classified into being roughly round or roughly elongated. It is possible, however, for two objects to have different outer contours, but still share interior features in common, particularly when those objects are from the same category. For example, an eel and a blowfish both have eyes, mouths, and other features that allow them to be recognized as fish. It may be impossible to fully separate the physical similarity between taxonomic coordinates, as shared features are part of the rationale for grouping objects into categories in the first place (Mirman et al., 2017), but there are certainly ways in which shape can be more assiduously controlled. For example, the overlap of specific features (e.g. eyes, claws, handles) rather than contours could be manipulated, and objective computational techniques could be leveraged to quantify similarity in shape.

The current set of results revealed differential verbal and nonverbal mechanisms as typical adults searched for thematic associates. This fits comfortably with a dual route account (Hurley et al., 2012; Hurley et al., 2018), in which humans are able to draw upon both verbal and nonverbal routes to access conceptual knowledge such as thematic associations.[2] The presence of two routes would provide robustness to a conceptual system: when one route is blocked by neurological damage, knowledge may instead be accessed *via* an alternate route. In the healthy brain, however, the neural pathways underling word and object recognition are richly interconnected (Mesulam, 1998; Binney et al., 2012; Zhang et al., 2016), which supports our ability to rapidly invoke imagery based on nouns (Reddy et al., 2010) and name pictures covertly (Ellis et al., 2006). To the extent that there are indeed separate routes for verbal and nonverbal knowledge, it is likely that typical adults in the current study flexibly drew upon both as they searched for thematic associates, as to do so would convey a conceptual advantage (Paivio, 1986). The current study design attempted to quarantine verbal from nonverbal stimuli, but we cannot dictate where those stimuli are processed internally after being transduced. The likelihood of "transmodal promiscuity" in the healthy brain limits our ability to attribute any given effect to a single network or pathway, even on within-platform trials.

---

[2]The current results can doubtlessly also be accommodated by a variety of unitary systems accounts, but these theoretical distinctions were not the central focus of the current study.

A logical next step would be to examine how search is affected by neurological syndromes where one route is selectively blocked, as occurs in disorders of language (aphasias) and object recognition (visual agnosias). Although the ventral streams of the object recognition and language networks run alongside one another in close proximity along the temporal lobe (Mesulam, 1998; Hickok and Poeppel, 2004), the critical nodes of each network may be differentially instantiated in each hemisphere. Corruption of word knowledge has been linked to left temporal lesions (Dronkers et al., 2004; Hurley et al., 2012), and inability to recognize objects has been linked to right temporal lesions (Butler et al., 2009; Hurley et al., 2018). One may thus predict temporal lobe damage to differentially affect either verbal or nonverbal search, depending on which hemisphere is most compromised.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Cleveland State University IRB. The participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

RH designed the paradigm, tested participants, analyzed data, prepared figures, and wrote the manuscript. JS analyzed data, prepared figures, and edited the manuscript. KN tested participants, analyzed data, prepared figures, and edited the manuscript. BL analyzed data, prepared figures, and edited the manuscript. WH designed the paradigm, prepared figures, and edited the manuscript. MS designed the paradigm, prepared figures, and edited the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcomm.2021.654575/full#supplementary-material

REFERENCES

Bajo, M.-T. (1988). Semantic Facilitation with Pictures and Words. *J. Exp. Psychol. Learn. Mem. Cogn.* 14, 579–589. doi:10.1037/0278-7393.14.4.579

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., et al. (2007). The English Lexicon Project. *Behav. Res. Methods* 39, 445–459. doi:10.3758/bf03193014

Bentin, S., Mccarthy, G., and Wood, C. C. (1985). Event-related Potentials, Lexical Decision and Semantic Priming. *Electroencephalography Clin. Neurophysiol.* 60, 343–355. doi:10.1016/0013-4694(85)90008-2

Bichot, N. P., Rossi, A. F., and Desimone, R. (2005). Parallel and Serial Neural Mechanisms for Visual Search in Macaque Area V4. *Science* 308, 529–534. doi:10.1126/science.1109676

Binney, R. J., Parker, G. J. M., and Lambon Ralph, M. A. (2012). Convergent Connectivity and Graded Specialization in the Rostral Human Temporal Lobe as Revealed by Diffusion-Weighted Imaging Probabilistic Tractography. *J. Cogn. Neurosci.* 24, 1998–2014. doi:10.1162/jocn_a_00263

Briihl, D., and Inhoff, A. W. (1995). Integrating Information across Fixations during reading: The Use of Orthographic Bodies and of Exterior Letters. *J. Exp. Psychol. Learn. Mem. Cogn.* 21, 55–67. doi:10.1037/0278-7393.21.1.55

Butler, C. R., Brambati, S. M., Miller, B. L., and Gorno-Tempini, M.-L. (2009). The Neural Correlates of Verbal and Nonverbal Semantic Processing Deficits in Neurodegenerative Disease. *Cogn. Behav. Neurol.* 22, 73–80. doi:10.1097/wnn.0b013e318197925d

Cave, K. R., and Wolfe, J. M. (1990). Modeling the Role of Parallel Processing in Visual Search. *Cogn. Psychol.* 22, 225–271. doi:10.1016/0010-0285(90)90017-x

Collins, A. M., and Loftus, E. F. (1975). A Spreading-Activation Theory of Semantic Processing. *Psychol. Rev.* 82, 407–428. doi:10.1037/0033-295x.82.6.407

Cooper, R. M. (1974). The Control of Eye Fixation by the Meaning of Spoken Language. *Cogn. Psychol.* 6, 84–107. doi:10.1016/0010-0285(74)90005-x

Dahan, D., Magnuson, J. S., and Tanenhaus, M. K. (2001). Time Course of Frequency Effects in Spoken-word Recognition: Evidence from Eye Movements. *Cogn. Psychol.* 42, 317–367. doi:10.1006/cogp.2001.0750

Dahan, D., and Tanenhaus, M. K. (2005). Looking at the Rope when Looking for the Snake: Conceptually Mediated Eye Movements during Spoken-word Recognition. *Psychon. Bull. Rev.* 12, 453–459. doi:10.3758/bf03193787

Donderi, D. C., and Mcfadden, S. (2005). Compressed File Length Predicts Search Time and Errors on Visual Displays. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* 49, 1649–1652. doi:10.1177/154193120504901732

Dronkers, N. F., Wilkins, D. P., Van Valin, R. D., Jr., Redfern, B. B., and Jaeger, J. J. (2004). Lesion Analysis of the Brain Areas Involved in Language Comprehension. *Cognition* 92, 145–177. doi:10.1016/j.cognition.2003.11.002

Ellis, A. W., and Young, A. W. (1988). *Human Cognitive Neuropsychology*. Hove U K ; Hillsdale (USA): L: Erlbaum Associates Publishers.

Ellis, A. W., Burani, C., Izura, C., Bromiley, A., and Venneri, A. (2006). Traces of Vocabulary Acquisition in the Brain: Evidence from covert Object Naming. *NeuroImage* 33, 958–968. doi:10.1016/j.neuroimage.2006.07.040

Estes, Z., Golonka, S., and Jones, L. L. (2011). "Thematic Thinking," in *The Psychology of Learning and Motivation: Advances in Research and Theory*. Editors B. H. Ross and B. H. Ross (San Diego, CA, US: Elsevier Academic Press), 54, 249–294. doi:10.1016/b978-0-12-385527-5.00008-5

Fluharty, M., Jentzsch, I., Spitschan, M., and Vishwanath, D. (2016). Eye Fixation during Multiple Object Attention Is Based on a Representation of Discrete Spatial Foci. *Sci. Rep.* 6, 31832. doi:10.1038/srep31832

Friedman, A., and Bourne, L. E. (1976). Encoding the Levels of Information in Pictures and Words. *J. Exp. Psychol. Gen.* 105, 169–190. doi:10.1037/0096-3445.105.2.169

Goto, K., Bond, A. B., Burks, M., and Kamil, A. C. (2014). Visual Search and Attention in Blue Jays (*Cyanocitta cristata*): Associative Cuing and Sequential Priming. *J. Exp. Psychol. Anim. Learn. Cogn.* 40, 185–194. doi:10.1037/xan0000019

Henderson, J. M., Williams, C. C., Castelhano, M. S., and Falk, R. J. (2003). Eye Movements and Picture Processing during Recognition. *Perception & Psychophysics* 65, 725–734. doi:10.3758/bf03194809

Hervet, G., Guérard, K., Tremblay, S., and Chtourou, M. S. (2011). Is Banner Blindness Genuine? Eye Tracking Internet Text Advertising. *Appl. Cognit. Psychol.* 25, 708–716. doi:10.1002/acp.1742

Hickok, G., and Poeppel, D. (2004). Dorsal and Ventral Streams: a Framework for Understanding Aspects of the Functional Anatomy of Language. *Cognition* 92, 67–99. doi:10.1016/j.cognition.2003.10.011

Huettig, F., and Altmann, G. T. M. (2007). Visual-shape Competition during Language-Mediated Attention Is Based on Lexical Input and Not Modulated by Contextual Appropriateness. *Vis. Cogn.* 15, 985–1018. doi:10.1080/13506280601130875

Huettig, F., and Altmann, G. T. M. (2005). Word Meaning and the Control of Eye Fixation: Semantic Competitor Effects and the Visual World Paradigm. *Cognition* 96, B23–B32. doi:10.1016/j.cognition.2004.10.003

Huettig, F., and Mcqueen, J. M. (2007). The Tug of War between Phonological, Semantic and Shape Information in Language-Mediated Visual Search. *J. Mem. Lang.* 57, 460–482. doi:10.1016/j.jml.2007.02.001

Hurley, R. S., Mesulam, M.-M., Sridhar, J., Rogalski, E. J., and Thompson, C. K. (2018). A Nonverbal Route to Conceptual Knowledge Involving the Right Anterior Temporal Lobe. *Neuropsychologia* 117, 92–101. doi:10.1016/j.neuropsychologia.2018.05.019

Hurley, R. S., Paller, K. A., Rogalski, E. J., and Mesulam, M. M. (2012). Neural Mechanisms of Object Naming and Word Comprehension in Primary Progressive Aphasia. *J. Neurosci.* 32, 4848–4855. doi:10.1523/jneurosci.5984-11.2012

Hutchison, K. A. (2003). Is Semantic Priming Due to Association Strength or Feature Overlap? A Microanalytic Review. *Psychon. Bull. Rev.* 10, 785–813. doi:10.3758/bf03196544

Inhoff, A. W., and Briihl, D. (1991). Semantic Processing of Unattended Text during Selective reading: How the Eyes See it. *Perception & Psychophysics* 49, 289–294. doi:10.3758/bf03214312

Iordanescu, L., Grabowecky, M., and Suzuki, S. (2011). Object-based Auditory Facilitation of Visual Search for Pictures and Words with Frequent and Rare Targets. *Acta Psychologica* 137, 252–259. doi:10.1016/j.actpsy.2010.07.017

Kalénine, S., Mirman, D., Middleton, E. L., and Buxbaum, L. J. (2012). Temporal Dynamics of Activation of Thematic and Functional Knowledge during Conceptual Processing of Manipulable Artifacts. *J. Exp. Psychol. Learn. Mem. Cogn.* 38, 1274–1295. doi:10.1037/a0027626

Kliegl, R., Grabner, E., Rolfs, M., and Engbert, R. (2004). Length, Frequency, and Predictability Effects of Words on Eye Movements in reading. *Eur. J. Cogn. Psychol.* 16, 262–284. doi:10.1080/09541440340000213

Kuisma, J., Simola, J., Uusitalo, L., and Öörni, A. (2010). The Effects of Animation and Format on the Perception and Memory of Online Advertising. *J. Interactive Marketing* 24, 269–282. doi:10.1016/j.intmar.2010.07.002

Lewis, G. A., Poeppel, D., and Murphy, G. L. (2015). The Neural Bases of Taxonomic and Thematic Conceptual Relations: An MEG Study. *Neuropsychologia* 68, 176–189. doi:10.1016/j.neuropsychologia.2015.01.011

Lohse, G. L. (1997). Consumer Eye Movement Patterns on Yellow Pages Advertising. *J. Advertising* 26, 61–73. doi:10.1080/00913367.1997.10673518

Lund, K., and Burgess, C. (1996). Producing High-Dimensional Semantic Spaces from Lexical Co-occurrence. *Behav. Res. Methods Instr. Comput.* 28, 203–208. doi:10.3758/bf03204766

Masson, M. E. J. (1995). A Distributed Memory Model of Semantic Priming. *J. Exp. Psychol. Learn. Mem. Cogn.* 21, 3–23. doi:10.1037/0278-7393.21.1.3

Mesulam, M. M. (1998). From Sensation to Cognition. *Brain* 121 (Pt 6), 1013–1052. doi:10.1093/brain/121.6.1013

Miellet, S., and Sparrow, L. (2004). Phonological Codes Are Assembled before Word Fixation: Evidence from Boundary Paradigm in Sentence reading. *Brain Lang.* 90, 299–310. doi:10.1016/s0093-934x(03)00442-5

Mirman, D., and Graziano, K. M. (2012a). Damage to Temporo-Parietal Cortex Decreases Incidental Activation of Thematic Relations during Spoken Word Comprehension. *Neuropsychologia* 50, 1990–1997. doi:10.1016/j.neuropsychologia.2012.04.024

Mirman, D., and Graziano, K. M. (2012b). Individual Differences in the Strength of Taxonomic versus Thematic Relations. *J. Exp. Psychol. Gen.* 141, 601–609. doi:10.1037/a0026451

Mirman, D., Landrigan, J.-F., and Britt, A. E. (2017). Taxonomic and Thematic Semantic Systems. *Psychol. Bull.* 143, 499–520. doi:10.1037/bul0000092

Nelson, D. L., Reed, V. S., and Mcevoy, C. L. (1977). Learning to Order Pictures and Words: A Model of Sensory and Semantic Encoding. *J. Exp. Psychol. Hum. Learn. Mem.* 3, 485–497. doi:10.1037/0278-7393.3.5.485

Nelson, D. L., Reed, V. S., and Walling, J. R. (1976). Pictorial Superiority Effect. *J. Exp. Psychol. Hum. Learn. Mem.* 2, 523–528. doi:10.1037/0278-7393.2.5.523

Nelson, W. W., and Loftus, G. R. (1980). The Functional Visual Field during Picture Viewing. *J. Exp. Psychol. Hum. Learn. Mem.* 6, 391–399. doi:10.1037/0278-7393.6.4.391

Norvig, P. (2012). English Letter Frequency Counts: Mayzner Revisited [Online]. Available: http://norvig.com/mayzner.html (Accessed.

Paivio, A., and Begg, I. (1974). Pictures and Words in Visual Search. *Mem. Cogn.* 2, 515–521. doi:10.3758/bf03196914

Paivio, A. (1986). *Mental Representations : A Dual Coding Approach.* New York: Oxford University Press.

Parker, R. E. (1978). Picture Processing during Recognition. *J. Exp. Psychol. Hum. Perception Perform.* 4, 284–293. doi:10.1037/0096-1523.4.2.284

Pieters, R., and Wedel, M. (2004). Attention Capture and Transfer in Advertising: Brand, Pictorial, and Text-Size Effects. *J. Marketing* 68, 36–50. doi:10.1509/jmkg.68.2.36.27794

Pollatsek, A., Raney, G. E., Lagasse, L., and Rayner, K. (1993). The Use of Information below Fixation in reading and in Visual Search. *Can. J. Exp. Psychology/Revue canadienne de Psychol. expérimentale* 47, 179–200. doi:10.1037/h0078824

Posner, M. I., Snyder, C. R., and Davidson, B. J. (1980). Attention and the Detection of Signals. *J. Exp. Psychol. Gen.* 109, 160–174. doi:10.1037/0096-3445.109.2.160

Proklova, D., Kaiser, D., and Peelen, M. V. (2016). Disentangling Representations of Object Shape and Object Category in Human Visual Cortex: The Animate-Inanimate Distinction. *J. Cogn. Neurosci.* 28, 680–692. doi:10.1162/jocn_a_00924

Rayner, K. (1998). Eye Movements in reading and Information Processing: 20 Years of Research. *Psychol. Bull.* 124, 372–422. doi:10.1037/0033-2909.124.3.372

Rayner, K., and Fisher, D. L. (1987). Letter Processing during Eye Fixations in Visual Search. *Perception & Psychophysics* 42, 87–100. doi:10.3758/bf03211517

Rayner, K. (2009). The 35th Sir Frederick Bartlett Lecture: Eye Movements and Attention in reading, Scene Perception, and Visual Search. *Q. J. Exp. Psychol.* 62, 1457–1506. doi:10.1080/17470210902816461

Reddy, L., Tsuchiya, N., and Serre, T. (2010). Reading the Mind's Eye: Decoding Category Information during Mental Imagery. *NeuroImage* 50, 818–825. doi:10.1016/j.neuroimage.2009.11.084

Rossell, S. L., Price, C. J., and Nobre, A. C. (2003). The Anatomy and Time Course of Semantic Priming Investigated by fMRI and ERPs. *Neuropsychologia* 41, 550–564. doi:10.1016/s0028-3932(02)00181-1

Rossion, B., and Pourtois, G. (2004). Revisiting Snodgrass and Vanderwart's Object Pictorial Set: the Role of Surface Detail in Basic-Level Object Recognition. *Perception* 33, 217–236. doi:10.1068/p5117

Schmidt, J., and Zelinsky, G. J. (2009). Short Article: Search Guidance Is Proportional to the Categorical Specificity of a Target Cue. *Q. J. Exp. Psychol.* 62, 1904–1914. doi:10.1080/17470210902853530

Schriefers, H., Meyer, A. S., and Levelt, W. J. M. (1990). Exploring the Time Course of Lexical Access in Language Production: Picture-word Interference Studies. *J. Mem. Lang.* 29, 86–102. doi:10.1016/0749-596x(90)90011-n

Seckin, M., Mesulam, M.-M., Rademaker, A. W., Voss, J. L., Weintraub, S., Rogalski, E. J., et al. (2016a). Eye Movements as Probes of Lexico-Semantic Processing in a Patient with Primary Progressive Aphasia. *Neurocase* 22, 65–75. doi:10.1080/13554794.2015.1045523

Seckin, M., Mesulam, M.-M., Voss, J. L., Huang, W., Rogalski, E. J., and Hurley, R. S. (2016b). Am I Looking at a Cat or a Dog? Gaze in the Semantic Variant of Primary Progressive Aphasia Is Subject to Excessive Taxonomic Capture. *J. Neurolinguist.* 37, 68–81. doi:10.1016/j.jneuroling.2015.09.003

Slattery, T. J., and Rayner, K. (2010). The Influence of Text Legibility on Eye Movements during reading. *Appl. Cognit. Psychol.* 24, 1129–1148. doi:10.1002/acp.1623

Smith, D. T., Ball, K., and Ellison, A. (2014). Covert Visual Search within and beyond the Effective Oculomotor Range. *Vis. Res.* 95, 11–17. doi:10.1016/j.visres.2013.12.003

Soanes, C., and Stevenson, A. (2004). *Concise Oxford English Dictionary*. Oxford University Press.

Sorensen, D. W., and Bailey, K. G. D. (2007). Object Perception, Attention, and Memory (OPAM) 2006 Conference Report. *Vis. Cogn.* 15, 69–123. doi:10.1080/13506280600975486

Sperber, R. D., Mccauley, C., Ragain, R. D., and Weil, C. M. (1979). Semantic Priming Effects on Picture and Word Processing. *Mem. Cogn.* 7, 339–345. doi:10.3758/bf03196937

Stenberg, G., Radeborg, K., and Hedman, L. R. (1995). The Picture Superiority Effect in a Cross-Modality Recognition Task. *Mem. Cogn.* 23, 425–441. doi:10.3758/bf03197244

Thierry, G., and Price, C. J. (2006). Dissociating Verbal and Nonverbal Conceptual Processing in the Human Brain. *J. Cogn. Neurosci.* 18, 1018–1028. doi:10.1162/jocn.2006.18.6.1018

Thorpe, S. J., Gegenfurtner, K. R., Fabre-Thorpe, M., and Bülthoff, H. H. (2001). Detection of Animals in Natural Images Using Far Peripheral Vision. *Eur. J. Neurosci.* 14, 869–876. doi:10.1046/j.0953-816x.2001.01717.x

Treisman, A. M., and Gelade, G. (1980). A Feature-Integration Theory of Attention. *Cogn. Psychol.* 12, 97–136. doi:10.1016/0010-0285(80)90005-5

Walther, D., and Koch, C. (2006). Modeling Attention to Salient Proto-Objects. *Neural Networks* 19, 1395–1407. doi:10.1016/j.neunet.2006.10.001

Wheatley, T., Weisberg, J., Beauchamp, M. S., and Martin, A. (2005). Automatic Priming of Semantically Related Words Reduces Activity in the Fusiform Gyrus. *J. Cogn. Neurosci.* 17, 1871–1885. doi:10.1162/089892905775008689

White, S. J. (2008). Eye Movement Control during reading: Effects of Word Frequency and Orthographic Familiarity. *J. Exp. Psychol. Hum. Perception Perform.* 34, 205–223. doi:10.1037/0096-1523.34.1.205

Wolfe, J. M., Horowitz, T. S., Kenner, N., Hyle, M., and Vasan, N. (2004). How Fast Can You Change Your Mind? the Speed of Top-Down Guidance in Visual Search. *Vis. Res.* 44, 1411–1426. doi:10.1016/j.visres.2003.11.024

Yee, E., Huffstetler, S., and Thompson-Schill, S. L. (2011). Function Follows Form: Activation of Shape and Function Features during Object Identification. *J. Exp. Psychol. Gen.* 140, 348–363. doi:10.1037/a0022840

Yee, E., Overton, E., and Thompson-Schill, S. L. (2009). Looking for Meaning: Eye Movements Are Sensitive to Overlapping Semantic Features, Not Association. *Psychon. Bull. Rev.* 16, 869–874. doi:10.3758/pbr.16.5.869

Zhang, W., Wang, J., Fan, L., Zhang, Y., Fox, P. T., Eickhoff, S. B., et al. (2016). Functional Organization of the Fusiform Gyrus Revealed with Connectivity Profiles. *Hum. Brain Mapp.* 37, 3003–3016. doi:10.1002/hbm.23222

Zim, H. S. (1966). *Codes and Secret Writing*. Scholastic Book Services.