Title
The Role of Preview and Incremental Delivery on Visual Search

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 34(34)

ISSN
1069-7977

Authors
Chiu, Eric
Spivey, Michael

Publication Date
2012

Peer reviewed

The Role of Preview and Incremental Delivery on Visual Search

Eric M. Chiu (echiu@ucmerced.edu)
Michael J. Spivey (spivey@ucmerced.edu)
Cognitive and Information Sciences, 5200 North Lake Road,
Merced, CA 95343 USA

Abstract

Recent studies show that visual search often involves a combination of both parallel and serial search strategies. Consequently, computational models and theoretical accounts of visual search processing have evolved from traditional parallel or serial descriptions to a continuum from "efficient" to "inefficient." In our first experiments (1a & 1b), we demonstrate with various conditions that search efficiency does not increase with *simultaneous* delivery of target features in a conjunction-search task. In the second experiment, we explore effects of *incremental* non-linguistic information delivery and discover improvement of search efficiency. We find a facilitatory effect when non-linguistic visual delivery of target features is concurrent with the visual display onset, but not when the target features are delivered prior to display onset. The results support an interactive account of visual perception that explains linguistic and non-linguistic mediation of visual search as chiefly due to the incrementality of target feature delivery once search has begun.

Keywords: visual search, incremental, conjunction, efficient

Introduction

The present study is part of a research program that explores the degree to which the incremental processing of spoken words in a full sentence can interact with concurrent visual search processes.

Traditionally, two contrasting perspectives have plagued the field of attention in visual search. The *serial-processing* perspective claims that observers allocate complete attentional resources discretely and wholly to individual objects, one at a time (Treisman & Gelade, 1980; Treisman, 1988). Conversely, biased competition has been found to be mediated by neural mechanisms in the extrastriate visual cortex, which forms a persuasive line of reasoning not in favor of the serial-processing perspective but for the *parallel-processing* perspective, which claims attention is better characterized as a function of partially active representations of objects simultaneously contending for probabilistic mappings onto motor output (Desimone & Duncan, 1995, Reynolds & Desimone, 2001; Desimone, 1998). Single-feature visual search has been demonstrated to be relatively unaffected by the number of distractors, often inducing a perceptual "pop-out" effect. In contrast, two-feature conjunction-searches typically produce a linear increase in reaction time (RT) as the number of distractors increase. However, as we will demonstrate these apparent dichotomous perspectives may not be from two contrasting fundamentals but rather a product of a single process better described as a graded enhancement of feature salience and supported by observations of improvement in visual search tasks (Olds, Cowan, & Jolicoeur, 2000a, 2000b, 2000c).

Olds and colleagues (2000a, 2000b, 2000c) observed, in a series of experiments, some facilitatory effects as a result of the very brief duration when search displays had only single-feature distractors. Although observers' responses were not as fast as with pure "pop-out" displays, Olds and colleagues (2000a, 2000b, 2000c) illustrated a graded improvement of search efficiency, by presenting single-feature visual search pop-out displays for very brief durations (in some conditions less than 100 milliseconds) before transitioning them to conjunction-search displays. Findings like this "search assistance," along with signal detection theory analyses of visual search data (Eckstein, 1998), and a lack of a bimodal search efficiency distribution (Wolfe, 1998), has replaced the serial-parallel dichotomy account with a continuum of search efficiency (e.g., Nakayama & Joseph, 1998).

Work by Spivey, Tyler, Eberhard, and Tanenhaus (2001) illustrates a different kind of "search assistance" phenomenon. Observers in an *Audio/Visual Concurrent* (A/V-concurrent) condition, where the conjunction-search display is presented concurrently with target identity via auditory linguistic queries (e.g. "Is there a red vertical?"), showed dramatically improved search efficiency compared to an *Auditory-First* control condition, where the same spoken query of target identity was provided prior to visual display onset. The findings suggest that upon hearing the first-mentioned adjective in the spoken query, visual attention is able to begin the search with only that feature, thus initiating the process more efficiency, resembling a single-feature search. Upon hearing the second adjective, several hundred milliseconds later, observers can then quickly find the target among the now attended subset of objects. Additionally, Reali, Spivey, Tyler, and Terranova (2006) implemented quantitative localist attractor simulations to extend the generalizability of the improvement in visual search efficiency when the identity of the conjunction target is delivered incrementally via a spoken target query while the stimulus display is visible, rather than prior to stimulus onset.

Subsequently, Gibson, Eberhard, and Bryant (2005) found with faster speech (4.8 syllables/second vs. 3.0 syllables/second) the A/V-concurrent condition no longer provides an enhanced efficiency effect on conjunction-search tasks, indicating that improvement in visual search efficiency is affected by speech rate.

Though more recently, experiments by Jones, Kaschak, and Boot (2011) used eye-tracking to examine an alternative view to one that proposes search efficiency is increased due to language enhancing perceptual processing. Jones and colleagues (2011) observed patterns of eye movements suggesting increased efficiency with concurrent speech was not likely due to linguistic enhancement of perceptual processes but instead delaying the onset of target-seeking eye movements. They contend the findings by Gibson et al. (2005) are better explained by this "preview" of search display (when observers are presented with the search display prior to being notified of the target object's identity) because slower speech provides observers with more search display viewing time, which provides additional information about potential target locations independently of the information conveyed by auditory linguistic speech stream.

The purpose of the present study was to, first, examine the role of preview of search display on visual processing and to, second, further understand exactly how language comprehension and visual search interact in real-time.

Experiment 1a

In this experiment, we utilized visual cues to deliver simultaneously a two-feature target identity in a conjunction-search task.

Method

In this experiment we utilized three SOA, stimulus onset asynchrony, conditions (0-ms, 350-ms, and 750-ms) when identifying the target object. Participants were either presented with the target identifying visual cue simultaneously with the search display (0-ms SOA) or with either a 350-ms or 750-ms delay after onset of search display. All three SOAs appeared equally and randomly.

Participants One hundred and fifty-seven University of California, Merced undergraduate students received course credit for participating in this experiment. Participants who were unable to perform the task with an accuracy of 80% or better were removed from the analysis. Twenty-four participants did not meet this requirement thus were removed from the analysis. Additionally, all incorrect responses and trials with RTs greater than 2.5 interquartile ranges (IQR) from the median were also omitted (IQR was used for data culling because of it's superior resistance to the influence of outliers).

Stimuli and Procedure Each stimulus bar subtended 2.8° X 0.4° of visual angle and neighboring bars were separated from one another by an average of 2.0° of visual angle. Target identifying visual cues were either red or green horizontal bars that appeared at the top and bottom of the search display or were red or green vertical bars that appeared on the left and right of the search display. Dimensions of the visual cues were designed to resemble the dimensions of the stimulus objects but four times larger.

The first block was referred to as the "practice" block, consisting of 32 trials, and was followed by an experimental block with 96 trials. Participants were instructed to respond to each display as quickly and accurately as possible by pressing the labeled "YES" button on the keyboard if the target was present in the display and the labeled "NO" button if it was absent.

The target object was present or absent in half of the trials. Moreover, we utilized four set sizes of objects (5, 10, 15, and 20), which appeared equally and randomly. Given two target features (color: red or green, and orientation: vertical or horizontal) four unique targets appeared equally and randomly throughout the trials. The duration of the entire experiment was approximately fifteen minutes. Two 20" Apple iMacs were used to run the experiment. The experiment was programmed and executed using Mathwork's MATLAB software.

Results and Discussion

In this experiment we demonstrate with various conditions that search efficiency does not increase in a conjunction-search task when target features are delivered simultaneously, despite having time to preview the search display. The RT-by-set-size functions for target-present trials (filled symbols) are shown in Figure 1 and Figure 2 for target-absent trials (open symbols) in the three SOA conditions, 0-ms (circles), 350-ms (diamonds), and 750-ms (triangles). We should note at this time that RT's were recorded from display onset, irrespective of condition, until a response was made. Next to each graph line is the best-fit linear equation and the proportion of variance accounted for (r^2). Error bars indicate standard error of the mean. In the 0-ms SOA control condition, the RT-by-set-size function was highly linear in both target-present, $r^2 = .994$, and target-absent trials, $r^2 = .984$, as typically observed in standard conjunction-search tasks. Similarly, the RT-by-set-size functions for the 350-ms and 750-ms SOA conditions were highly linear in target-present trials, $r^2 = .925$ and $r^2 = .992$, and target-absent trials, $r^2 = .977$ and $r^2 = .961$, respectively.

Since our primary interest is to assess the effects of preview on visual search efficiency, analysis in this experiment compared the 350-ms and 750-ms SOA conditions to the 0-ms SOA control condition. Overall mean RTs, as well as y-intercepts, were significantly slower in the 350-ms and 750-ms SOA conditions because delivery of target identity was delayed by 350-ms and 750-ms, respectively, relative to the 0-ms SOA control condition for both target-present, $t(132) = 2.38$, $p = .017$, and $t(132) = 8.21$, $p < .001$, and for target-absent, $t(132) = 4.05$, $p < .001$, $t(132) = 9.31$, $p < .001$, trials. Similar to previous observations, mean accuracy was 94.7% for all three conditions (Spivey et al., 2001; Reali et al., 2006).

For all experiments in this report the most important analysis is in comparison of the slopes of functions relating RT to set size. This slope value is an indicator of how efficient the search process is, that is, how much it resembles a serial process where each new distractor object increases RT by a sizeable fixed duration, or how much it resembles a parallel process where each new distractor object increases RT by little or no amount. The slopes of the RT-by-set-size functions reveal that 350-ms and 750-ms SOA conditions did not produce more efficient visual search compared with the 0-ms SOA control conditions (see fig. 1 & 2). Contrary to findings by Jones et al. (2011) an analysis revealed slopes for the 350-ms and 750-ms SOA conditions compared to the 0-ms SOA control condition were not significantly different for target-present trials (22.4 ms/item & 20.5 ms/item vs. 19.6 ms/item), $t(132) = 0.61$, $p = .543$, and $t(132) = 0.21$, $p = .835$, and target-absent trials (37.0-ms/item & 35.7 ms/item vs. 41.9 ms/item), $t(132) = -0.99$, $p = .323$, and $t(132) = -1.26$, $p = .207$.

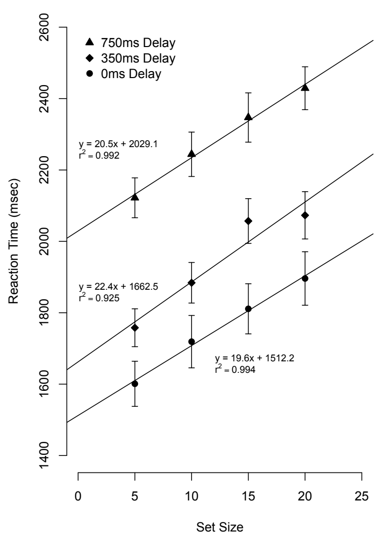

Figure 1. Results from Experiment 1a shown separately for target-present trials.

Similar to Spivey et al. (2001) and Reali et al. (2006), we found a near 2:1 ratio between target-absent and -present trials in all three conditions (37.0-ms/item vs. 22.4 ms/item for 0-ms SOA, 35.7 ms/item vs. 20.5 ms/item for 350-ms SOA, and 41.9 ms/item vs. 19.6 ms/item for 750-ms SOA). This 2:1 ratio between target-absent and -present trials has been regarded as consistent with a standard serial search account.

The results of this experiment indicate that simply delivering target identity simultaneously in a conjunction-search task with a variety of SOAs so that observers are allowed preview time does not substantially affected search efficiency. The results observed in all three conditions are of the type that are traditionally interpreted as consistent with the construction of a conjunction template of the target object followed by a serial process whereby discretely comparing each display object with the target template (Treisman & Gelade, 1980).

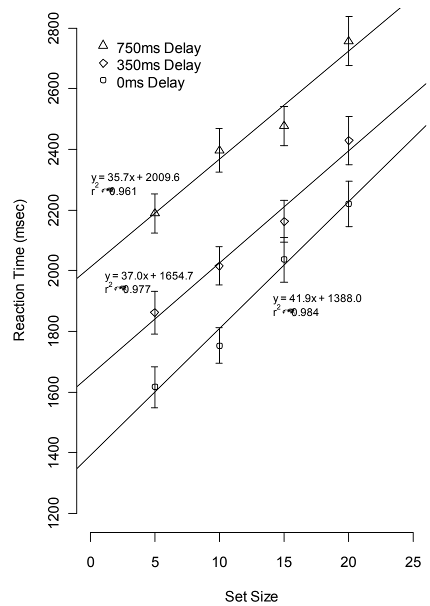

Figure 2. Results from Experiment 1a shown separately for target-absent trials.

Experiment 1b

In this experiment we extended the methods in Experiment 1a to first, mimic the duration (1500-ms) of the auditory linguistic query, which identified the target object in previous work by Spivey et al. (2001) and to, secondly, explore the effects of a relatively long preview duration of search display on visual search processing.

Methods

The method in the experiment follows that of Experiment 1a with the exception that only two SOAs (0-ms and 1500-ms) were used for the target identifying visual cue.

Participants Fifty-nine University of California, Merced undergraduate students received course credit for participating in this experiment. Five participants were unable to perform the task with an accuracy of 80% or better and were subsequently removed from the analysis. As with Experiment 1a, all incorrect responses and trials with RTs greater than 2.5 IQRs from the median were also omitted.

Stimuli and Procedure The same stimuli and target identifying visual cues from Experiment 1a were used in this experiment. Participants were presented with both SOAs equally and randomly in a within-subjects experimental design. The same testing apparatuses and software were used in this experiment as the last.

Results and Discussion

As with Experiment 1a, we continue to demonstrate with a slightly different condition that search efficiency does not increase with simultaneous delivery of target feature in a conjunction-search task, despite having time to preview the search display. Figure 3 shows the RT-by-set-size functions for target-present trials (filled symbols) and target-absent trials (open symbols) in the 0-ms SOA (triangles) and 1500-ms SOA (circles). In the 0-ms SOA control condition, the RT-by-set-size function was highly linear in both target-present, r^2 = .995, and target-absent trials, r^2 = .979, as typically observed in standard conjunction-search tasks. Similarly, the RT-by-set-size functions for the 1500-ms SOA condition was highly linear in target-present trials, r^2 = .975, and target-absent trials, r^2 = .958.

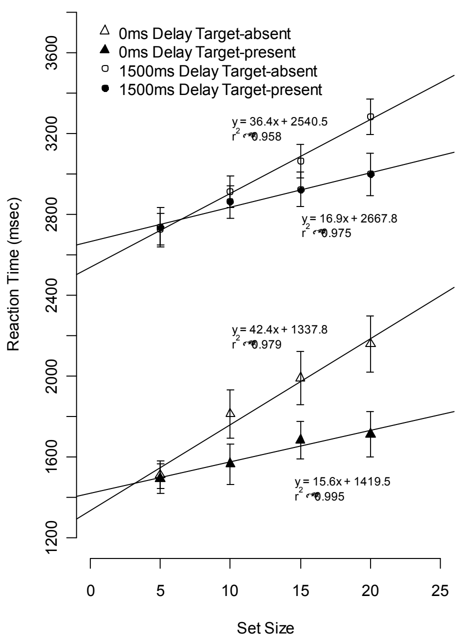

Figure 3. Results for Experiment 1b shown separately for target-present and –absent trials.

Overall mean RTs, as well as y-intercepts, were significantly slower in the 1500-ms SOA condition because delivery of target identity was delayed by 1500-ms relative to the 0-ms SOA control condition for both target-present, $t(53) = -3.05$, $p = .002$, and target-absent, $t(53) = -3.06$, $p < .002$, trials. Similar to previous observations, mean accuracy was 94.5% for both conditions.

The slopes of the RT-by-set-size functions reveal that the 1500-ms SOA condition did not produce more efficient visual search compared with the 0-ms SOA control conditions (see fig. 3). An analysis revealed slopes for the 1500-ms SOA condition compared to the 0-ms SOA control condition were not significantly different for target-present trials (16.9 ms/item vs. 15.6 ms/item), $t(53) = 0.22$, $p = .825$, and target-absent trials (36.4 ms/item vs. 42.4 ms/item), $t(53) = -0.85$, $p = .398$.

Consistent with Experiment 1a, we found a near 2:1 ratio between target-absent and -present trials in both conditions (36.4 ms/item vs. 16.9 ms/item and 42.4 ms/item vs. 15.6 ms/item).

The results of this experiment continue to indicate that simply delivering target identity simultaneously in a conjunction-search task with a relatively long SOA (1500-ms). From Spivey et al. (2001) participants were unable to effectively utilize a noteworthy preview to improve search efficiency. Although a 1500-ms SOA mimics the overall duration of linguistic query it fails to simulate the incremental delivery of information characteristic of speech.

Experiment 2

In this experiment, we explore effects of incremental non-linguistic information delivery on visual search processing by visually replicating the temporal characteristics of the auditory linguistic query that was used to identify the target object in previous work by Spivey et al. (2001).

Methods

In this experiment, to simulate the auditory-first and A/V-concurrent condition in Spivey et al. (2001), we utilized two slightly different conditions. A *cue-first* condition similar to the auditory-first condition, delivered target identity incrementally via a visual cue prior to display onset, and a *cue-concurrent* condition similar to the A/V-concurrent condition, delivered target identity incrementally via the identical visual cue but concurrently with display onset. In Spivey et al. (2001) all participants failed to report experiencing any difference in display onset timing, thus auditory-first and A/V-concurrent conditions appeared randomly in a mixed trial design. Since in our experiment the difference between timing of display onset for cue-first and cue-concurrent trials was much more apparent, due to the unimodal nature of the task, we opted for a blocked trial design.

Participants Forty-six University of California, Merced undergraduate students received course credit for participating in this experiment. Eight participants were unable to perform the task with a minimum accuracy of 80% and were subsequently removed from the analysis. As with the previous experiments, all incorrect responses and trials with RTs greater than 2.5 IQRs from the median were also omitted from the analysis.

Stimuli and Procedure Stimulus objects were identical to experiments 1a and 1b. In order to visually simulate the incremental information delivery of the spoken query (e.g., "Is there a red vertical?" 500-ms to utter the first feature color, "red" or "green," and 1000-ms to utter the second feature orientation, "vertical" or "horizontal") in Spivey et al. (2001), target identifying visual cues began as all red or all green horizontal and vertical bars that appeared on all sides (top, bottom, left, and right) of the search display for 500-ms to identify the color of the target. To identify the orientation of the target, the visual cue then transitioned to grey horizontal or vertical bars that appeared either at the top and bottom or the left and right of the search display, respectively, for 1000-ms before disappearing. Dimensions of the visual cues were identical to the previous experiments.

Prior to participating in the experimental blocks observers participated in two practice blocks (one of each cue-first and cue-concurrent) consisting of 32 total trials and was followed by two experimental blocks with 64 trials each for a total of 128 trials. One experimental block contained cue-first trials only and the other contained cue-concurrent trials only. The order of the experimental blocks (cue-first first or cue-concurrent first) was randomly assigned to participants, each order was used equally. Participants were instructed to respond to each display as quickly and accurately as possible by pressing the labeled "YES" button on the keyboard if the target was present in the display and the labeled "NO" button if it was absent.

The target object was present or absent in half of the trials. Furthermore, we utilized four set sizes of objects (5, 10, 15, and 20), which appeared equally and randomly. Given two target features (color: red or green, and orientation: vertical or horizontal) four unique targets appeared equally and randomly throughout the trials. The duration of the entire experiment was approximately 20 minutes. Two 20" Apple iMacs were used to run the experiment. The experiment was programmed and executed using Mathwork's MATLAB software.

Results and Discussion

In this experiment we demonstrated a facilitatory effect when visual non-linguistic delivery of target features is presented concurrently with the visual display onset, but not when the target features are delivered prior to display onset. Figure 4 shows the RT-by-set-size functions for target-present trials (filled symbols) and target-absent trials (open symbols) in the cue-first (triangles) and cue-concurrent (circles) conditions. In both target-present and -absent trials the RT-by-set-size function was linear for both the cue-concurrent condition, $r^2 = .768$, $r^2 = .962$, respectively, and the cue-first condition, $r^2 = .314$, $r^2 = .698$, respectively, which is typically observed in standard conjunction-search tasks. Overall mean RT, as well as y-intercepts, were significantly slower in the cue-concurrent condition because delivery of target identity was delayed by 1500-ms relative to the cue-first control condition for both target-present, $t(37) = 4.49$, $p < .001$, and target-absent, $t(37) = -4.32$, $p < .001$, trials. Mean accuracy was 94.0% for both conditions.

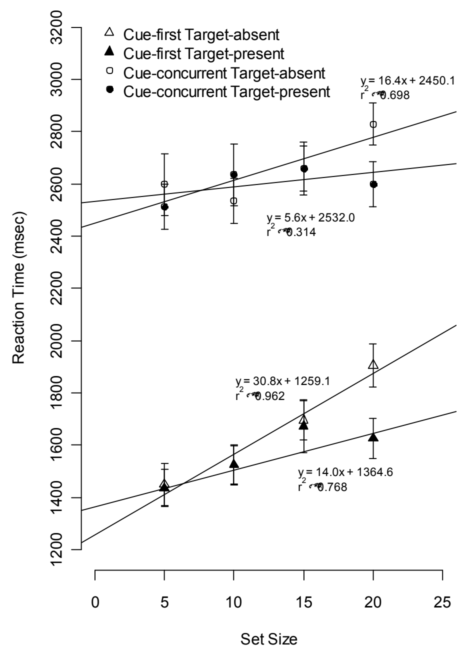

Figure 4. Results from Experiment 2 shown separately for target-present and –absent trials.

The slopes of the RT-by-set-size functions reveal that the cue-concurrent conditions produced more efficient visual search compared with the cue-first control conditions (see fig. 4). An analysis revealed slopes for the cue-concurrent condition compared to the cue-first control condition were significantly different for target-present trials (5.6 ms/item vs. 14.0-ms/item), $t(37) = -2.77$, $p = .010$, and target-absent trials (16.4 ms/item vs. 30.8 ms/item), $t(37) = -2.75$, $p = .006$. Furthermore, we found a near 2:1 ratio between target-absent and -present trials in both cue-concurrent conditions (16.4 ms/item vs. 5.6 ms/item) and cue-first conditions (30.8 ms/item vs. 14.0-ms/item), regarded as consistent with a standard serial search account.

The results of this experiment indicate that visual non-linguistic delivery of target features presented incrementally and concurrently with the visual display onset has a facilitatory effect on visual search efficiency, but not when the target features are delivered prior to display onset. The results observed in the cue-first condition are of the type that are traditionally interpreted as consistent with the construction of a conjunction template of the target object followed by a serial process of sequentially comparing each display object with the target template (Treisman & Gelade, 1980). Conversely, the results in the cue-concurrent condition, which simply involved shifting the relative timing of display onset relative to target identity cue, are more consistent with a parallel or "partial parallel" (Maioli, Benaglio, Siri, Sosta, & Cappa, 2001) search process, which is observed in the similarly shallower slopes in the RT-by-set-size functions.

General Discussion

In a series of experiments we made strides toward understanding exactly how language comprehension and visual search interact in real-time. We demonstrated with various conditions that search efficiency does not increase with simultaneous delivery of target features in a conjunction-search task despite relatively lengthy previews of search display, 1500-ms in some conditions (Experiment 1a & 1b). We then explored the effects of incremental non-linguistic information delivery by visually simulating auditory linguistic queries and discovered an improvement of search efficiency where facilitatory effect only occurred when visual non-linguistic delivery of target features was concurrent with the visual display onset, and not when the target features were delivered prior to display onset.

In conclusion, our findings suggest that it is the incremental nature of target delivery (whether via speech perception or visual perception) that allows the visual search process to begin when only a single feature of the target identity has been heard. When the initial feature is identified the search proceeds in an efficient nearly-parallel fashion so when the second adjective is presented, a substantial amount of the target identification process has already been completed, and as a result the presence of multiple distractors is less disruptive. These results support an interactive account of visual perception that explains linguistic and non-linguistic mediation of visual search as chiefly due to the incrementality of target feature delivery once search has begun. Future research on understanding exactly how language comprehension and visual search interact in real-time will benefit greatly from the development of further experimental tests such as this.

Acknowledgments

We are grateful to Andreas Kolling for help with the MATLAB code for the experiments. Additional thanks to Monica Yanez, Markie Johnson, Norma Cardona, Mauricio Cifuentes and Courtney Griffin-Oliver for help with collecting data.

References

Desimone, R. (1998). Visual attention mediated by biased competition in extrastriate visual cortex. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *353*(1373), 1245.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*(1), 193–222.

Eckstein, M. P. (1998). The lower visual search efficiency for conjunctions is due to noise and not serial attention processing, *Psychological Science, 9*, 111-118.

Gibson, B. S., Eberhard, K. M., & Bryant, T. A. (2005). Linguistically mediated visual search: The critical role of speech rate. *Psychonomic Bulletin and Review*, *12*(2), 276.

Jones, J. J., Kaschak, M. P., & Boot, W. R. (2011). Language mediated visual search: The role of display preview. *Cognitive Science Proceedings*, 2739-2744.

Maioli, C., Benaglio, I., Siri, S., Sosta, K., & Cappa, S. (2001). The integration of parallel and serial processing mechanisms in visual search: Evidence from eye movement recording. *European Journal of Neuroscience. 13*(2), 364-372.

Nakayama, K., & Joseph, J. S. (1998). Attention, pattern recognition, and pop-out in visual search. *The attentive brain*, 279–298.

Olds, E. S., Cowan, W. B., & Jolicoeur, P. (2000a). Partial orientation pop-out helps difficult search for orientation. *Perception & psychophysics*, *62*(7), 1341–1347.

Olds, E. S., Cowan, W. B., & Jolicoeur, P. (2000b). The time-course of pop-out search. *Vision Research*, *40*(8), 891–912.

Olds, E. S., Cowan, W. B., & Jolicoeur, P. (2000c). Tracking visual search over space and time. *Psychonomic Bulletin and Review*, *7*(2), 292–300.

Reali, F., Spivey, M. J., Tyler, M. J., & Terranova, J. (2006). Inefficient conjunction-search made efficient by concurrent spoken delivery of target identity. *Perception and Psychophysics*, *68*(6), 959.

Reynolds, J., & Desimone, R. (2001). Neural mechanisms of attentional selection. *Visual attention and cortical circuits (Braun J, Koch C, Davis JL, eds)*, 121–136.

Spivey, M. J., Tyler, M. J., Eberhard, K. M., & Tanenhaus, M. K. (2001). Linguistically mediated visual search. *Psychological Science*, *12*(4), 282-286.

Treisman, A., & Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review*, *95*(1), 15–48.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive psychology*, *12*(1), 97–136.

Wolfe, J. M. (1998). What can 1 million trials tell us about visual search? *Psychological Science, 9*, 33-39.