

Working memory load predicts visual search efficiency: Evidence from a novel pupillary response paradigm

Nada Attar¹ · Matthew H. Schneps¹ · Marc Pomplun¹

Published online: 27 April 2016
© Psychonomic Society, Inc. 2016

Abstract An observer's pupil dilates and constricts in response to variables such as ambient and focal luminance, cognitive effort, the emotional stimulus content, and working memory load. The pupil's memory load response is of particular interest, as it might be used for estimating observers' memory load while they are performing a complex task, without adding an interruptive and confounding memory test to the protocol. One important task in which working memory's involvement is still being debated is visual search, and indeed a previous experiment by Porter, Troscianko, and Gilchrist (*Quarterly Journal of Experimental Psychology*, 60, 211–229, 2007) analyzed observers' pupil sizes during search to study this issue. These authors found that pupil size increased over the course of the search, and they attributed this finding to accumulating working memory load. However, since the pupil response is slow and does not depend on memory load alone, this conclusion is rather speculative. In the present study, we estimated working memory load in visual search during the presentation of intermittent fixation screens, thought to induce a low, stable level of arousal and cognitive effort. Using standard visual search and control tasks, we showed that this paradigm reduces the influence of non-memory-related factors on pupil size. Furthermore, we found an early increase in working memory load to be associated with more efficient search, indicating a significant role of working memory in the search process.

Keywords Working memory · Visual search · Eye movements · Reaction time analyses

Our conscious visual experience and the performance of everyday visual tasks are strongly determined by the interaction of visual attention and working memory. Both of these cognitive mechanisms have been extensively studied (for recent reviews, see Carrasco, 2011, and Baddeley, 2012). One of the most common tasks that we arguably perform thousands of times every day is visual search, which is also one of the major paradigms for studying visual attention (Eckstein, 2011). Researchers have used various visual search paradigms to gain insight into attentional selection in the visual system. In a typical visual search task, participants are asked to report whether a visually distinctive target object is present among a set of distractors in a given scene. If the target and distractors have similar visual characteristics, observers have to sequentially attend to the search items to find the target or determine its absence. Many models of visual search have been proposed, aimed at explaining the role of visual attention (e.g., Cave & Wolfe, 1990; Treisman & Gelade, 1980; Wolfe, 1994).

Despite this research effort, the role of visual working memory in visual search is still being debated. Obviously, for successful task completion, the information about the search target needs to be kept in memory during search. This working memory content guides visual attention during the search process, both in artificial search displays (e.g., Cave & Wolfe, 1990) and in natural images (e.g., Pomplun, 2006). Interestingly, memory content guides search even if it is irrelevant to the search task (Downing, 2000). It is unclear, however, whether and to what extent working memory is involved in the search process beyond target memorization. Most importantly, working memory could be used to remember those

✉ Nada Attar
nattar@cs.umb.edu

¹ Department of Computer Science, University of Massachusetts at Boston, Boston, MA 02125, USA

items that have already been inspected, in order to avoid another, redundant inspection of the same item. If working memory can be utilized in this way without introducing significant delay, it could substantially facilitate search.

The results of many studies have suggested a significant role of memory in visual search (e.g., Baddeley, 1986; Logie, 1995; Luck & Vogel, 1997; Peterson, Kramer, Wang, Irwin, & McCarley, 2001; Phillips, 1974), and other studies have shown that memory can assist attention during the search task (e.g., Klein & MacInnes, 1999). However, Horowitz and Wolfe (1998) claimed that there is no memory involvement in visual search. In their experiment, they asked subjects to search for the presence of a letter “T” among letters “L” in either a random or a static condition. In both conditions, a new frame occurred every 111 ms, during the final 28 ms of which the letters were individually masked. In the random condition all letters were randomly relocated in each frame, but the letters remained constant in the static condition. The static condition should have allowed subjects to keep track of the inspected items, in contrast to the dynamic condition, in which the target could have appeared at any previously scanned location. Surprisingly, search efficiency did not differ between the two conditions, as indicated by equally steep slopes for response time (RT) as a function of set size. From these results, Horowitz and Wolfe inferred that visual search has no memory.

However, Peterson, Kramer, Wang, Irwin, and McCarley (2001) questioned the conclusion that the mechanism responsible for guiding attention during search has no memory for the locations that have already been visited. In their study, they found that the pattern of object revisitations did not fit the prediction of a memoryless search model. This finding suggests that observers can keep track of what they have previously inspected during visual search. A series of further studies supported this claim by providing evidence for memory involvement in visual search and showing that working memory can influence search performance on a number of levels (see Gilchrist & Harvey, 2000; Shore & Klein, 2000).

Another approach to investigating the role of working memory in visual search is the study of the relationship between an observer’s working memory capacity and his or her visual search performance. If search performance were directly correlated with working memory capacity, this would support the hypothesis that the use of working memory facilitates visual search. Unfortunately, whereas some researchers have found such a relationship (e.g., Anderson et al., 2013; Emrich et al., 2009), others have concluded that it does not exist (e.g., Kane, Poole, Tuholski, & Engle, 2006).

It is important to note that the direct assessment of working memory load through memory retrieval tasks would interfere with the search task. Consequently, the results of such experiments would be difficult to interpret. An unobtrusive way of estimating working memory capacity and use during an

ongoing task has been discovered through electroencephalography (EEG) research (e.g., Emrich et al., 2009; Vogel & Machizawa, 2004). For example, Emrich et al. (2009) investigated an electrophysiological marker of visual working memory encoding and maintenance, termed the *contralateral delay activity* (CDA), in the context of visual search. They found the variations in CDA amplitude to be correlated with both visual working memory capacity and visual search efficiency, indicating a direct relationship between the latter two variables.

Although the electrophysiological approach allows for visual working memory load estimation without requiring a secondary task, it does require the availability and setup of EEG equipment, which adds considerable cost and effort to any given study. To overcome this problem, a different method of estimating working memory load, which also does not require the introduction of an additional task, can be applied. This technique is based on pupillometry, whose relationship with cognitive load was first explored in the 1960s. For example, Hess and Polt (1964) measured the pupillary responses of people engaged in performing mental arithmetic problems of increasing complexity. They found that the level of difficulty of the problem correlated positively with pupil diameter (see also Ahern & Beatty, 1979; Kahneman, 1973). Extending their research, Kahneman and Beatty (1966) related the variation in pupil dilation to memory load. Their participants were asked to remember series of digits orally presented by the examiner at the rate of one digit per second. After a short pause, participants were asked to retell the verbal digits. The results showed that during the encoding of each digit being presented, there was an increase in pupil size, providing evidence of a gradual increase in memory load. Furthermore, when reproducing the digits aloud, pupil size decreased for each individual digit, reflecting a parallel decrease in memory load. In trials with a critical number of digits involved, the pupil size would remain large during the entire procedure, indicating a somewhat sustained effect of increased memory load. Many subsequent studies have further illustrated this correspondence between pupillary dilation and working memory load (e.g., Ahern & Beatty, 1979; Karatekin, Couperus, & Marcus, 2004; Nuthmann & van der Meer, 2005; Stanners, Coulter, Sweet, & Murphy, 1979).

Porter, Troscianko, and Gilchrist (2007) used pupillometry to study visual working memory load during visual search. Their visual search and counting task experiments manipulated search difficulty by varying the number of distractors as well as the heterogeneity of the distractors, and the dilatory patterns were compared between the two tasks. The results indicated an almost constant, large pupil size during the counting task. In contrast, during search, pupil size increased from the start of the trial onward, which the authors interpreted as showing increasing working memory load as the search progressed. However, although this conclusion is plausible,

it was not entirely assured by their data, due to the many factors that influence pupil size. For example, the amount of effort that an observer spent finding the target could have been a contributing factor. Furthermore, the pupil response to changes in working memory load is slow (e.g., Beatty & Lucero-Wagoner, 2000). It is thus possible that the working memory load did not increase during the search trial but was constant, and that the observed gradual increase in pupil size was due only to the delayed pupil response.

To overcome this problem, in the present study we modified the search paradigm used by Porter et al. (2007). At regular time intervals, the search display was replaced with an intermittent blank screen showing only a central fixation marker. Subjects were asked to fixate the marker until the search display returned, and then to resume their search. We hypothesized that measuring pupil size during these fixation intervals should lead to more reliable estimates of working memory load than measuring pupil size during search intervals. Our reasoning was that during the fixation intervals, the only mental effort required of the subjects was the memorization of the target object features and any information about the current search progress that they were keeping track of. While fixating the marker, subjects did not spend any effort inspecting the search items and did not encounter any target-relevant information, and therefore these two factors could not influence pupil size. It is also sensible to assume that, during these intervals, the absence of any cognitive task except the maintenance of working memory would induce no systematic variation in arousal levels that could interfere with memory load assessment. In addition, the fixation intervals were relatively long (3 s), so that the pupil size could approximate its equilibrium value. We also ensured that the average display luminance remained constant throughout all screens used in the experiment. This requirement minimized light-induced changes in pupil size that would otherwise have added noise to our measurements.

In our experiments reported here, we used this paradigm to test whether the use of working memory beyond memorizing the search target features is important for efficient search. Such working memory use would be indicated by increasing memory load during the course of a search trial. Therefore, if this memory use facilitates search, we would expect to find a negative correlation of the increase in working memory load and the time taken to find the target. In other words, in those trials in which an observer used his or her working memory more strongly, leading to a greater increase in working memory load, we would expect that this memory use would facilitate search. This facilitation should, on average, lead to faster target detection, indicated by shorter RTs. Since visual search trials could vary greatly in their duration, it was a sensible approach to measure the increase in working memory load early during the trial, in order to collect a maximum amount of data during an experiment. Consequently, we computed the

amount of early memory load increase as the change in load between the first and second fixation screens in a trial. As we noted above, increasing working memory load tends to be accompanied by dilation of the observer's pupils. Therefore, we used as an indicator of working memory load increase the average pupil size during the second fixation display minus the one measured during the first fixation display. As a control, we also computed this value for the first and second search screens in each trial.

The correlation of the resulting load difference with the subject's RT in the same trial was then computed to test the hypothesis that a greater memory load increase would lead to more efficient search. Such an effect would be indicated by a negative correlation coefficient. Furthermore, if our new paradigm were effective at reducing the interference of other cognitive processes with the measurement of working memory load, this inverse correlation should be stronger when measuring pupil size during the fixation phases than during the search phases.

To gather baseline data for our paradigm, in Experiment 1 we used a modified visual search task that strictly required working memory encoding for its successful completion. In this task, the search items in each stimulus display were arranged randomly on four (invisible) circles, each of which included a random number of visually distinguishable target items. After subjects had indicated task completion, they were asked to report the number of targets on each circle. Clearly, to perform the task in an efficient manner, subjects needed to scan the display circle by circle and keep in memory the number of targets in each previously inspected circle. Consequently, their memory load was expected to increase steadily over time. According to our memory hypothesis, a faster increase in working memory load early in a trial should be associated with more efficient search, as long as the subjects searched in a systematic fashion. Therefore, besides analyzing the correlation between early changes in pupil size and RT, we also examined the subjects' search strategies in order to obtain baseline data for the effectiveness of our paradigm at estimating the moment-to-moment working memory load.

Subsequently, in Experiment 2, we employed a "standard" visual search task and collected the same type of behavioral and pupil data. We compared a memory and a no-memory condition to investigate the use of working memory during visual search. In the memory trials, the same search stimulus was displayed after each intermittent screen, whereas in the no-memory trials, a different search display was shown after each intermittent screen. The no-memory condition thus did not allow subjects to accumulate task-relevant information about the search items across search intervals. Comparing the correspondence between pupil size and RT between the tasks thus allowed us to study the impact of working memory use on search task performance. The results indicated that working memory plays an important role in this task and that

the new paradigm can provide insight into working memory use during various tasks.

Experiment 1

Method

Subjects Thirteen students at the University of Massachusetts at Boston were recruited for the experiment. All subjects were between the ages of 21 and 35 years old and right-handed, with normal or corrected-to-normal vision. Each subject was compensated \$10 for participating in the 1-h experiment.

Apparatus On a separate computer, eye movements were recorded using an SR Research EyeLink 1000 desktop system with a sampling frequency of 1000 Hz. After calibration, the average calibration error was 0.5° . Stimuli were presented on a 22-in. ViewSonic LCD monitor with a refresh rate of 75 Hz and a screen resolution of $1,024 \times 768$ pixels. All viewers sat at a distance of approximately 70 cm from the screen in a room with a dim light setup and used a chin rest to stabilize their head. Only the left eye was tracked.

Materials Sixty stimulus displays were generated using a MATLAB script. Each display consisted of 32 search items, which were Gabor patches with a diameter of 1° . These items were arranged randomly on four circles with radii of 5° , arranged in a 2×2 array. To avoid the overlapping and visual crowding of items, we set a distance of at least 3° between the centers of each item pair. Each stimulus was placed at the center of the screen and subtended a visual angle of $26^\circ \times 26^\circ$. Three to five of the Gabor patches were the designated search targets, indicated by their vertical or horizontal orientation. In each search display, it was randomly determined whether all targets had the same vertical or horizontal orientation. The distractors were oriented randomly with a minimum angular difference of 12° from both the vertical and horizontal orientations. The three, four, or five targets in each display were distributed randomly across the four circles, so that these circles did not necessarily contain the same numbers of targets. The intermittent fixation screen had a gray background of the same luminance as the search screens. Examples of these stimuli are shown in Fig. 1a.

Procedure After participants had signed consent forms approved by the University of Massachusetts Boston institutional review board (IRB) and read the instructions, a standard 9-point grid calibration and validation of the gaze recording were completed. Participants were instructed to look at the central fixation marker on the pretrial and intermittent screens and search for the targets on each stimulus screen, counting the number of targets in each circle. For each trial, the search

display was shown for 1 s and the intermittent screen for 3 s, followed by the same search display for 1 s and another fixation screen for 3 s, and so on. All trials started and ended with a fixation screen shown for 3 s.

Participants were to press the space bar when they had finished counting the targets in each circle. Once the space bar had been pressed or the maximum number of ten search displays had been reached, a new screen was displayed with four options for the potential numbers of targets in the different circles (see Fig. 1b). Participants were asked to choose the only option that showed the correct number of targets in each circle by pressing the corresponding button on a gamepad. For example, if the four circles contained two, one, zero, and one targets, respectively, then the option they should choose is 2, 1, 0, 1, where the order of the circles is described left-to-right and top-to-bottom. Subsequently, the next trial would start. Each subject performed four practice trials, followed by 60 experimental trials that were grouped into six blocks of ten trials. The same set of displays was used for all participants in an individually randomized order.

Results

In this experiment, we analyzed pupil size to investigate changes in working memory load during visual search. Furthermore, we analyzed eye movements to examine the participants' shifts of attention while they performed the search task. We specifically investigated the interaction between the pupil size and visual scanning strategies to explain differences in search efficiency across participants.

All correct trials in which the participants found the targets within fewer than three search displays were removed from the analysis. Even though we only used the pupil data from the first two fixation and search displays, potential target detections within these screens could have influenced the subjects' pupil sizes. Although it was almost impossible to complete the task in Experiment 1 within fewer than three search screens, this often occurred in Experiment 2. To ensure comparability of the results between the two experiments, we set the three-display limit for both of them. Furthermore, all trials that had an incorrect response were removed from the analysis, as well. These criteria limited the analysis to approximately 82 % of the trials in the entire experiment. The RT was measured from the time when the first search display appeared—that is, after the initial fixation screen until the time that the subject pressed the space bar, indicating that all targets had been counted. The time after the space bar being pressed and the next screen with the four options being shown was not included in the RT. The mean RT of those trials that were accepted according to the criteria was 25.57 s, with a standard deviation of 7.25 s.

For each subject, the correlation between (a) the difference in mean pupil size between the entire interval of the second intermittent fixation screen and the entire interval of the first

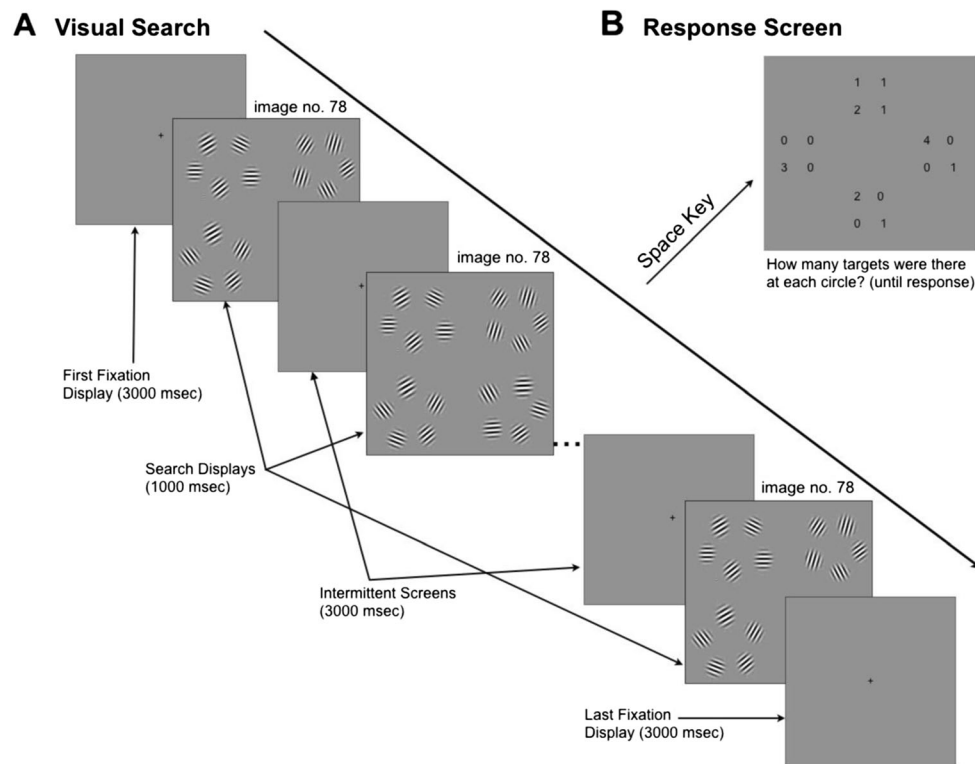


Fig. 1 Sample trial of Experiment 1, using intermittent fixation screens in a visual search task that included the counting of target objects. (a) Observers had to determine and remember the number of targets in each circle during search and disregard the distractors. Note that the actual displays contained eight objects per circle. After completing this task, subjects were to press the space bar on the keyboard, which would replace the search display with the multiple-choice response screen (b). This screen had four different response options appear at the top, bottom,

left, and right positions on the screen, each of which presented four numbers corresponding to the numbers of targets in the different circles. For example, if one of the options consisted of the numbers 2, 0, 0, and 1, as is shown in the option at the bottom of panel B, this meant that the upper-left circle had two targets, the lower-right circle had one target, and the remaining two circles had no targets. One of the four options indicated the correct four numbers of targets in the four circles, and subjects were to pick this option by pressing one of four buttons on a gamepad

intermittent fixation screen and (b) the RT was computed across all trials. Similarly, the correlation was also calculated between the pupil size difference between the second and first search screens and the RT. Pupil size was measured only during fixations, to minimize artifacts due to eye movements, and was taken as the number of pixels that the pupil covered in the EyeLink camera image. No further normalization of this variable was performed because we only investigated pupil size within individual subjects.

It should be noted that in video-based eyetracking the measured pupil size slightly varies with the gaze position, because of the varying angle between the pupil and the eyetracker camera. Although we had used a pupil calibration and normalization method (Pomplun & Sunkara, 2003) for our previously used EyeLink II system, we found that the system used in this study, the EyeLink 1000, provided significantly more robust pupil size measurements. In the present experiments, we found that the standard deviation in the average measured pupil size for fixation positions across the four screen quadrants during search was approximately 20 pixels, or 1.25 % of the mean pupil size. Given the overall variation in pupil size

(cf. Fig. 4 below), even a systematic bias of the pupil size data by eye movements could have only a small effect on the results. However, we should be aware that when comparing the results obtained through pupil size measurement during central fixation (fixation screens) and during free viewing (search screens), the latter may contain some additional noise.

The pupil size increase during intermittent screens showed a slightly negative correlation with RT ($r = -.085$), $t(12) = 1.410$, $p = .006$, whereas measurement during the search displays revealed a weak positive correlation ($r = .124$), $t(12) = 2.286$, $p = .005$. This difference in the correlations between the two methods of measurement was significant, $t(12) = 2.609$, $p = .023$.

Although we found the hypothesized negative correlation between pupil size increase and RT for the pupil measurements during the fixation screens, it was very weak. One possible explanation for this finding is that some subjects did not perform the task systematically, and therefore did not load their working memory during the initial screens as expected. To examine this possibility, we divided the participants into two groups based on their r value for the fixation screens, with

the cutting point set at the mean r value, which was $-.085$. Therefore, the first group, termed the “strong-correlation” group, consisting of six subjects, had correlations with $r < -.085$, with an average of $r = -.287$. The second group, named the “weak-correlation” group, had correlations with $r > -.085$, with an average of $r = .10$. We compared the gaze patterns in the search displays between the two groups in an attempt to understand the reason for the difference in correlations.

In the present task, for efficient performance, subjects had to process each circle in a serial pattern because they were asked to report the numbers of targets in the different individual circles. Examining their eye movements allowed us to determine how systematic a subject’s search strategy was. We analyzed the fixations during the first five search displays of each trial. Each fixation was assigned to the circle of Gabor patches whose center had the shortest Euclidean distance from it. Then, for each 1-s presentation of a search display, we determined the circle that received the most fixations and divided that number of fixations by the total number of fixations recorded during that display presentation. If subjects focused their search on one circle during each display presentation, which we assumed would reflect systematic search behavior, this variable would approach a value of 100 %. If, on the other hand, subjects showed erratic search patterns, the value could in the most extreme case fall to 25 %.

Figure 2 shows the separate results for the strong- and weak-correlation groups for each of the first five search displays in each trial. Participants in the weak-correlation group revealed lower values (mean = .67) than those in the high-correlation group (mean = .8), $t(11) = 3.16$, $p = .022$, indicating that they tended to fixate on several circles instead of one circle during each search display. This finding suggests that the weak-correlation group performed less systematic search, with fixations being distributed among more than one circle.

To further study the efficiency of observers’ search strategies, we considered that inefficient search performance can be indicated by a substantial proportion of eye movements revisiting one of the stimulus circles. A fixation was considered a revisit if it landed on a circle that had been visited before, with at least one fixation on a different circle occurring between these visits. These fixations were assumed to be made to recheck an already searched circle, which would be indicative of unsystematic search behavior.

As is shown in Fig. 3, the weak-correlation group had to return to search a circle whose items had already been examined almost once at every second and third search display, averaging at 0.69 revisits per display. The fixations for the strong-correlation group showed only 0.28 refixations, which was significantly lower than the value for the weak-correlation group, $t(11) = 2.83$, $p = .053$. Interestingly, the difference between the groups was extremely large for the second search display, but progressively diminished until it disappeared for the fifth display. This finding is most likely due to the

necessity of systematic search to revisit circles if some of the target numbers determined earlier could not be memorized. Figure 4 shows the correlations for one of the subjects in the strong-correlation group and illustrates the difference that it makes to measure pupil size during the search or during the fixation screens.

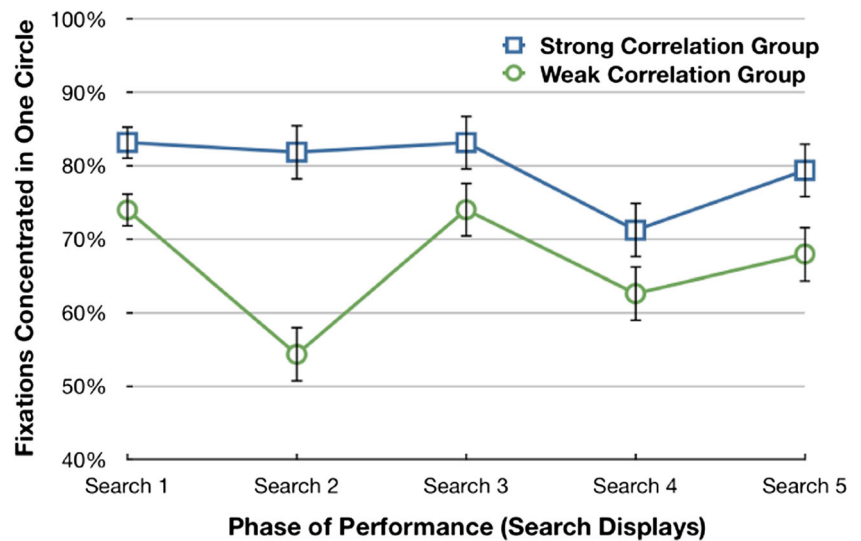
Discussion

We hypothesized a negative correlation between pupil size increases during the early intermittent screens and RTs, indicating that a greater working memory load increase led to faster searches. This hypothesis was supported by our pupil size analysis. Pupil dilation during the intermittent screens had a small but significant negative correlation with RT, whereas there was a small positive correlation between pupil dilation and RT during the search displays.

To determine whether our pupillary results were affected by the subjects’ search strategies, we divided the participants into two different groups based on the mean value of the correlation between pupil size in the first two intermittent screens and RT. We compared their gaze patterns during the search displays. Our results indicated that, in general, the group with a strong negative correlation between pupil dilation and RT tended to search the stimulus circles in a systematic sequence, with fewer revisits of previously inspected circles. These results gave further support to our hypothesis that, using the fixation screens, working memory load during the search task can be estimated with useful accuracy. Although a correlation of $r = -.29$ is still weak, it is important to note that no strong correlation could be expected for these variables. RTs in visual search tasks generally show great variance, because by chance the target object may be the first one inspected, or in the worst case, it could be the last one. Even when using efficient search strategies, the RT difference between the best and worst cases is immense. Even though the need for finding multiple targets in Experiment 1 reduced this variance, we still could not expect to precisely predict RT in individual trials on the basis of pupil size changes.

When measuring pupil size while subjects were actively searching the display, no such negative correlation was found, suggesting that these data do not seem to allow reliable estimates of working memory load. As we noted above, the additional noise in pupil size measurement during free viewing may have contributed, at least to a small extent, to this difference between the fixation and search screens as the basis of measurement. Having obtained these supporting baseline data for our method, we conducted Experiment 2, in which we applied the method to a more “regular” visual search task that did not explicitly demand memory use, except for memorizing the target features.

Fig. 2 Average percentages of fixations on the same circle in each of the first five search displays for all participants in each group. The strong-correlation group revealed more systematic serial search, with 72 %–83 % of the fixations landing on the same circle during each search display. In contrast, the values for the weak-correlation group ranged from 54 % to 74 %. Error bars show the standard errors of the means



Experiment 2

The search task in Experiment 2 can be considered a standard visual search task in which the subject had to report the identity of one target among a set of distractors. We introduced two experimental conditions—a memory task and a no-memory task. Each task was presented in separate blocks of trials. In the memory task, after each intermittent screen, the same search stimulus reappeared. Therefore, subjects could make use of any information that they had memorized from previous screens, such as the set of already-inspected items. In the no-memory task, a new search stimulus appeared after every intermittent fixation screen. Consequently, subjects did not need to memorize any information across search displays within the same trial, because this would be useless in the next search display.

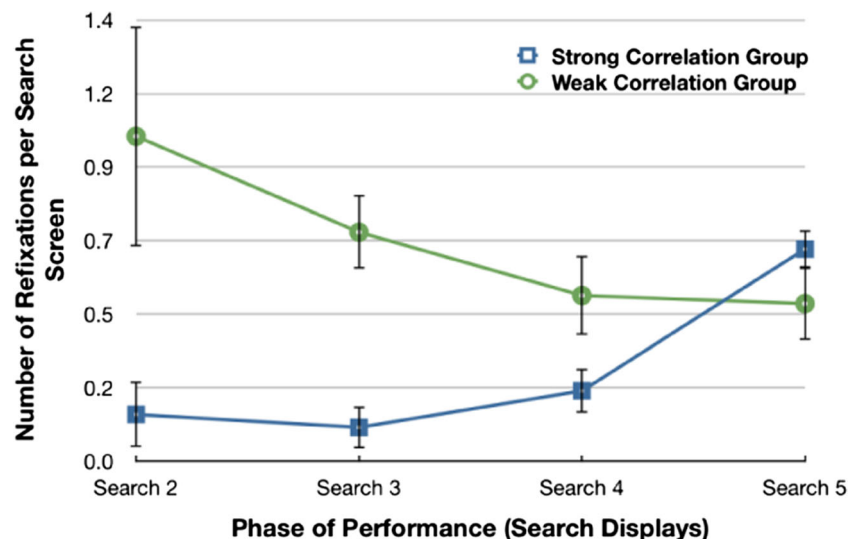
In the memory task, we hypothesized that the loading of working memory was an integral part of performing the search

task efficiently, as only a single search display was presented in each trial. As in Experiment 1, we expected this role of working memory to be revealed by a negative correlation between the pupil size increase between the first and second fixation screens and RT. The no-memory task served as a control condition in which no increasing memory load or its correlation with RT were expected. Once again, pupil measurement during the fixation screens was contrasted with pupil measurement during the search screens, to obtain further evidence for the suitability of our working memory load estimation paradigm.

Method

Subjects Thirteen students at the University of Massachusetts Boston were recruited for the experiment. All of the subjects were between the ages of 19 and 36 years old and right-handed,

Fig. 3 Results from Experiment 1, showing the average numbers of refixations per search display for each of the two subject groups. Inefficient search behavior by the weak-correlation group is indicated by many fixations revisiting previously inspected circles. Error bars show the standard errors of the means



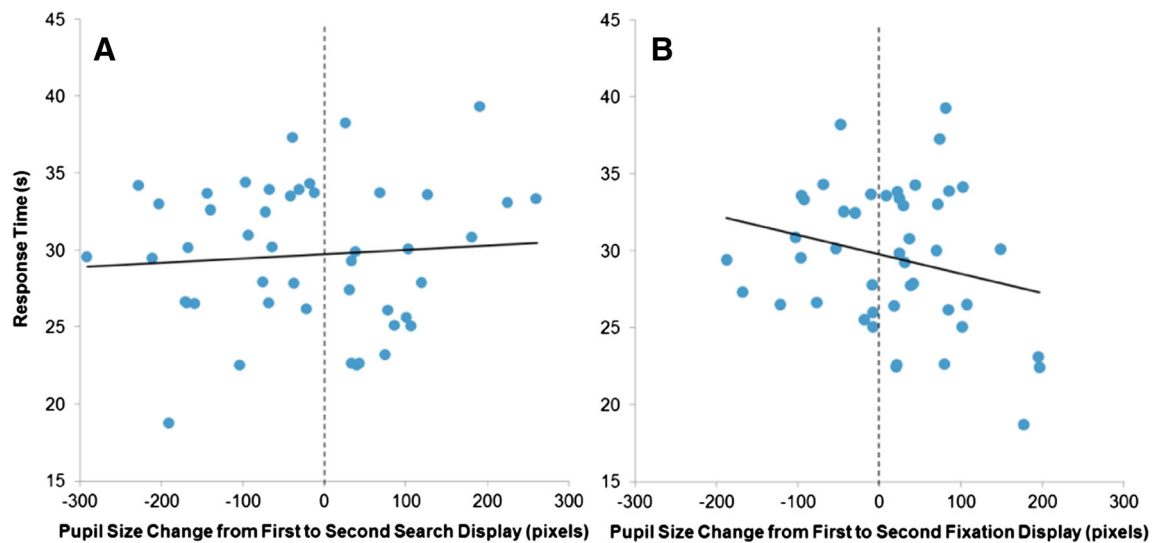


Fig. 4 Sample correlations between pupil size changes and response times for the same subject in Experiment 1 (Subject 6 from the strong-correlation group). Each marker represents one trial with a correct response, and straight lines indicate the results of linear regressions. (a)

A slightly positive correlation ($r = .07$) when pupil size was measured during search displays; (b) a weak negative correlation ($r = -.24$) when pupil size was measured during fixation displays

with normal or corrected-to-normal vision. Each participant was compensated \$10 for participating in the 1-h experiment.

Apparatus This was the same as in Experiment 1.

Materials A total of 440 stimulus displays was generated using a MATLAB script. Each display subtended a visual angle of $26^\circ \times 26^\circ$ at the center of the screen and contained a total of 31 distractors, which were Gabor patches with a radius of 1° . They were oriented randomly with angles differing from both the vertical and horizontal orientations by at least 12° . Furthermore, each display contained a single target oriented vertically or horizontally. To ensure that the objects did not overlap, we set a minimum distance of 3° of visual angle between their centers.

After we had randomized the orientation of the target in each display, we divided the displays into two groups, forming four blocks of memory trials using a total of 40 displays, and four blocks of no-memory trials with a total of 400 displays. Since all of the displays were randomly chosen, the displays in the no-memory condition differed within each trial, and thus a mix of horizontal and vertical targets could appear for the different search displays in a given no-memory trial. When a subject responded by pressing a key (“v” or “h”) to indicate whether the target was vertical or horizontal for the current search display, the response would be considered on the basis of the last image display that the participant had seen. In other words, the subject could press the response keys at any time during the trial, and only the orientation of the target in the last-presented search display before the user’s response was counted. The two task conditions, memory and no-memory, were administered to the subjects in an alternating order, with

ten trials per block. The intermittent screens showed a central fixation marker on a gray background matching the average luminance of the search displays. Examples of these stimuli are shown in Fig. 5.

Procedure After participants had signed consent forms approved by the University of Massachusetts at Boston IRB and read the instructions, a standard 9-point grid calibration and validation of the gaze recording were completed. Participants were instructed to look at the central fixation marker on the intermittent screens and search for the targets on each stimulus screen. They were asked to respond using two keyboard keys when they had detected a target at any time during the trial, with “v” indicating a vertical target object and “h” indicating a horizontal one. The search displays were shown for 1 s, and the intermittent screens for 3 s. Each trial continued until the target type was reported or until the search display had been shown ten times. All trials started and ended with a fixation screen for 3 s.

Results

The subjects in the memory and no-memory conditions, respectively, had 84 % and 86 % correct responses. In the same way as in Experiment 1, we considered only trials that had a correct response and in which the participant had found the target after the third search display. These criteria limited the analysis to approximately 50 % of the trials in each condition. For each subject, their mean pupil size was measured for the first two intermittent screens and the first two search displays. This was done separately for the memory and no-memory blocks. Those trials that we kept for analysis showed mean

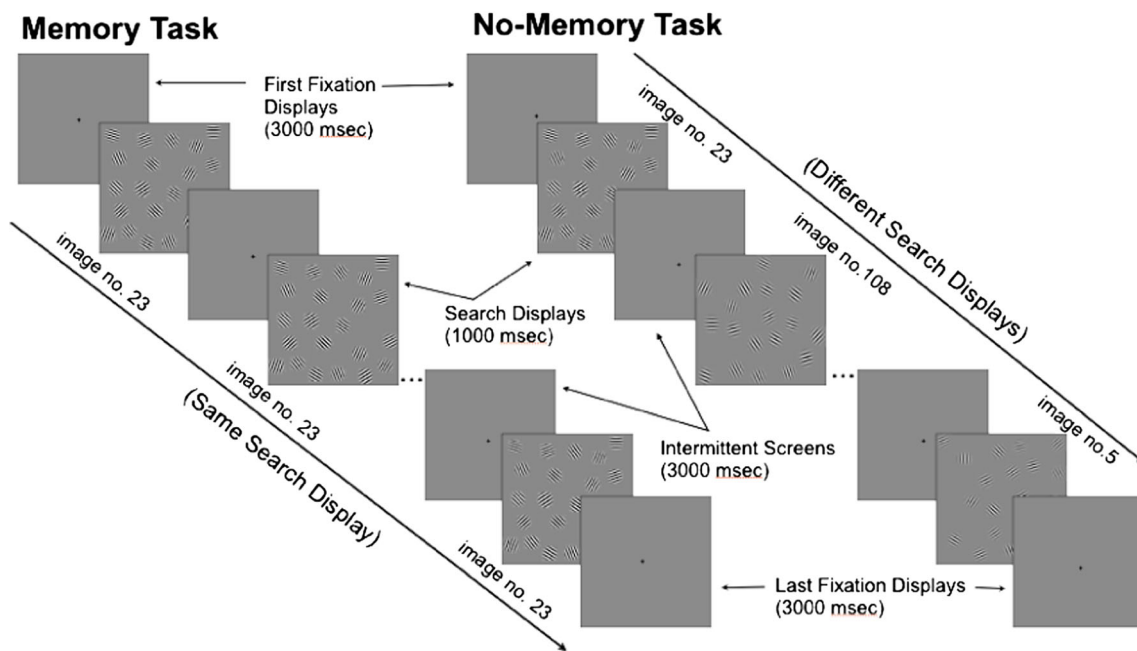


Fig. 5 Trial schematics for the two experimental conditions in Experiment 2. Both conditions used typical visual search displays, but the search process was interrupted by presenting intermittent screens for unbiased pupil size measurement. (Left) A trial in the memory condition,

with the same search display presented during the entire trial, (Right) A no-memory trial, with a different search display after each intermittent screen. Note that, as in Experiment 1, each display actually contained 32 Gabors, here one target and 31 distractors

RTs of 18.75 and 20.85 s in the memory and no-memory conditions, respectively, with standard deviations of 9.79 and 10.00 s. There was no statistical difference between these two RT values, $t(12) = 2.15$, $p = .054$. The numbers of fixations per search display, averaging at 4.25 and 4.28, respectively, during each of the first two displays, did not differ between the memory and no-memory conditions, $t(12) = 0.57$, $p = .58$.

Two factors were used in the data analysis. The first was the factor Phase, indicating whether pupil size was measured during the intermittent screens or search displays. The second factor was Task (memory or no-memory). The effects of these factors on the correlations between pupil size increase during the first two measurements and RT are shown in Fig. 6. A repeated measures two-way analysis of variance using these two factors indicated a marginal effect of task, $F(1, 12) = 3.87$, $p = .067$, and a significant effect of phase, $F(1, 12) = 4.93$, $p = .044$. The interaction between the two factors was also significant, $F(1, 12) = 8.77$, $p = .013$.

This result of the within-subjects analysis indicated that the difference in mean pupil size between the first two successive fixation screens in the memory condition mainly reflected changes in working memory load that occurred during the search interval between the screens. This difference was a significant predictor of the RT in the same trial, with an inverse correlation of approximately $r = -.22$, $t(12) = 4.37$, $p = .045$. We found no significant correlations for the intermittent

screens in the no-memory blocks, $t(12) = 1.46$, $p = .170$, or for the pupil size measurement during search phases in either the memory, $t(12) = 1.15$, $p = .122$, or the no-memory, $t(12) = 1.02$, $p = .182$, conditions.

The results indicate that during the memory blocks, greater pupil size increases in the intermittent fixation screens tended to be followed by shorter search times. Furthermore, the difference in mean pupil size between the first two search display presentations did not predict RTs in either the memory or no-

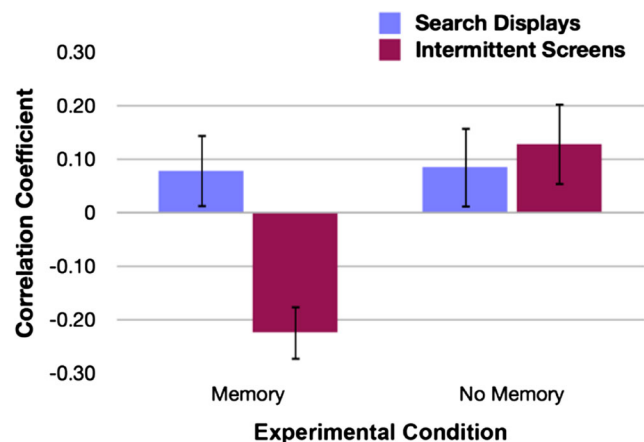


Fig. 6 Correlation between pupil size changes between the first two displays (either intermittent fixation screens or search displays) and the response time in the same trial for the memory and no-memory conditions in Experiment 2. Error bars indicate standard errors of the means

memory blocks. Figure 7 shows sample correlation plots for one of the subjects in Experiment 2.

Discussion

Surprisingly, the results of Experiment 2, which did not involve an explicit working memory requirement, revealed a clearer characterization of memory load effects than did Experiment 1, designed to require memory encoding and retrieval. In the “repeated displays” (memory task) condition of Experiment 2, in which working memory use could have facilitated search, a faster increase in memory load was correlated with better search performance. As in Experiment 1, the correlation was weak, which was expected due to the random variance in RT data. Since Experiment 2 required the detection

of only one target among 31 distractors, the influence of the randomly chosen target position on RTs was actually much stronger than in Experiment 1. Therefore, finding a significant correlation between pupil size increase and RT in this condition of Experiment 2 indicates an important role of memory in this task.

This correlation was only found when we measured during fixation screens, and disappeared when we measured during search screens. In the “changing displays” (no-memory task) condition, in which performance could not have benefited from working memory use, no significant correlations were found. This pattern of results supports the initial hypotheses: Subjects do tend to make extensive use of working memory in a standard visual search task, and this significantly improves their search performance.

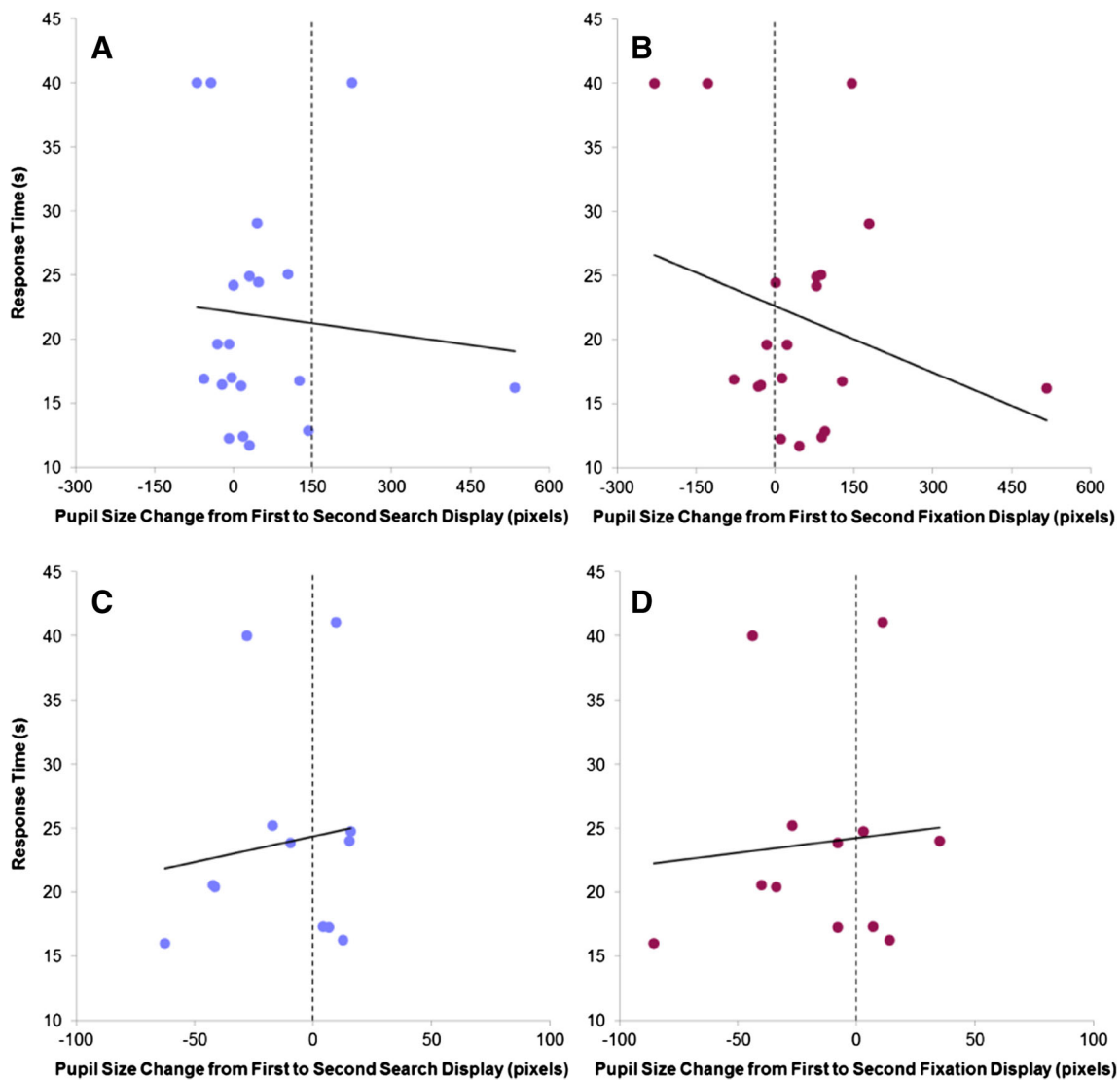


Fig. 7 Sample correlations between pupil size changes and response times for the same subject in Experiment 2 (Subject 1). Each marker represents one trial with a correct response and at least three search displays, and the straight lines indicate the results of linear regressions.

The upper row (a and b) shows the memory condition, and the lower row (c and d) the no-memory condition. The left column (a and c) refers to pupil size measurements during the search displays, and the right column (b and d) to pupil size measurements during the fixation displays

Conclusions

The Porter et al. (2007) study provided valuable new insights into cognitive load during visual search tasks. Their moment-to-moment analysis was the first to successfully use pupillometry in such a context. However, the influence of factors other than memory on pupil size did not allow for specific conclusions about the role of working memory in visual search. To enable such an investigation, we modified their approach by introducing a new paradigm that we hypothesized would better suit measuring memory load independently from the other factors that influence pupil size.

The present study had two main objectives. First, we aimed at testing the hypothesis that measuring pupil size during intermittent fixation screens, assumed to induce a low, stable level of arousal and cognitive effort, would allow a more reliable estimation of working memory load than measuring it during the actual task performance. Second, in our study we investigated the relationship between memory load increase and search performance. The pupil dilation data in Experiment 1, serving as a baseline, showed that the increase in pupil size during the first two intermittent screens had a greater negative correlation with RT than did that measured during search displays. Experiment 2 confirmed these results for a more common type of visual search task. By contrasting memory and no-memory conditions, this experiment provided further evidence for the feasibility of measuring working memory using the new experimental paradigm. Furthermore, the present results indicate that subjects make significant use of working memory during visual search, which improves their performance. This conclusion is consistent with previous studies that have proposed a close link between working memory use and search efficiency (e.g., Downing, 2000; Kristjánsson, 2000; Peterson et al., 2001; Shore & Klein, 2000).

We need to note that the reduction of eye movement noise in video-based pupil size measurement by having subjects fixate on a marker may have slightly contributed to the observed advantage of our proposed method. Furthermore, we should be aware of the fact that the introduction of the intermediate fixation screens altered the task to a certain extent. The method presented here can only be applied to tasks with typical RTs of at least a few seconds, which are the only tasks for which a time course of memory load would be of interest. The task manipulation could lead to changes in the cognitive task requirements and task performance. However, such changes should not substantially affect subjects' task-relevant memory use, and therefore the memory load measurement would still be relevant for the original task.

Our technique could be applied to a variety of other tasks to measure the working memory load that they induce, without requiring an explicit working memory test that could interfere with the performance of the main task. For example, the

paradigm could be used to investigate working memory use in the context of language comprehension or arithmetic problem solving. Such studies may lead to a more comprehensive understanding of the interaction of visual attention and working memory that is crucial for the performance of most everyday tasks and our conscious experience.

Author note Part of the study was supported by Grant No. R01 EY021802 from the National Institutes of Health to M.P. We thank the Saudi Arabian Cultural Mission, Ministry of Higher Education, for supporting the first author with a Ph. D. scholarship. The authors also thank the undergraduate students at the Visual Attention Laboratory at the University of Massachusetts for their help recruiting subjects.

References

- Ahern, S., & Beatty, J. (1979). Pupillary responses during information processing vary with scholastic aptitude test scores. *Science*, *205*, 1289–1292.
- Anderson, D. E., Vogel, E. K., & Awh, E. (2013). A common discrete resource for visual working memory and visual search. *Psychological Science*, *24*, 929–938. doi:10.1177/0956797612464380
- Baddeley, A. D. (1986). *Working memory*. Oxford, UK: Oxford University Press, Clarendon Press.
- Baddeley, A. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology*, *63*, 1–29. doi:10.1146/annurev-psych-120710-100422
- Beatty, J., & Lucero-Wagoner, B. (2000). The pupillary system. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of psychophysiology* (2nd ed., pp. 142–162). Cambridge, UK: Cambridge University Press.
- Carrasco, M. (2011). Visual attention: The past 25 years. *Vision Research*, *51*, 1484–1525. doi:10.1016/j.visres.2011.04.012
- Cave, K. R., & Wolfe, J. M. (1990). Modeling the role of parallel processing in visual search. *Cognitive Psychology*, *22*, 225–271. doi:10.1016/0010-0285(90)90017-X
- Downing, P. E. (2000). Interactions between visual working memory and selective attention. *Psychological Science*, *11*, 467–473. doi:10.1111/1467-9280.00290
- Eckstein, M. P. (2011). Visual search: A retrospective. *Journal of Vision*, *11*(5), 14:1–36. doi:10.1167/11.5.14
- Emrich, S. M., Al-Aidroos, N., Pratt, J., & Ferber, S. (2009). Visual search elicits the electrophysiological marker of visual working memory. *PLoS ONE*, *4*, e8042. doi:10.1371/journal.pone.0008042
- Gilchrist, I. D., & Harvey, M. (2000). Refixation frequency and memory mechanisms in visual search. *Current Biology*, *10*, 1209–1212. doi:10.1016/S0960-9822(00)00729-6
- Hess, E. H., & Polt, J. M. (1964). Pupil size in relation to mental activity during simple problem-solving. *Science*, *143*, 1190–1192.
- Horowitz, T. S., & Wolfe, J. M. (1998). Visual search has no memory. *Nature*, *357*, 575–577. doi:10.1038/29068
- Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice-Hall.
- Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. *Science*, *154*, 1583–1585.
- Kane, M. J., Poole, B. J., Tuholski, S. W., & Engle, R. W. (2006). Working memory capacity and the top-down control of visual search: Exploring the boundaries of “executive attention”. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 749–777. doi:10.1037/0278-7393.32.4.749

- Karatekin, C., Couperus, J. W., & Marcus, D. J. (2004). Attention allocation on the dual task paradigm as measured through behavioral and psychophysiological responses. *Psychophysiology*, *41*, 175–185.
- Klein, R. M., & MacInnes, W. J. (1999). Inhibition of return is a foraging facilitator in visual search. *Psychological Science*, *10*, 346–352. doi:10.1111/1467-9280.00166
- Kristjánsson, Á. (2000). In search of remembrance: Evidence for memory in visual search. *Psychological Science*, *11*, 328–332. doi:10.1111/1467-9280.00265
- Logie, R. H. (1995). *Visuo-spatial working memory*. Hove, UK: Erlbaum.
- Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*, 279–281. doi:10.1038/36846
- Nuthmann, A., & van der Meer, E. (2005). Time's arrow and pupillary response. *Psychophysiology*, *42*, 306–317.
- Peterson, M. S., Kramer, A. F., Wang, R. X. F., Irwin, D. E., & McCarley, J. S. (2001). Visual search has memory. *Psychological Science*, *12*, 287–292. doi:10.1111/1467-9280.00353
- Phillips, W. A. (1974). On the distinction between sensory storage and short-term visual memory. *Perception & Psychophysics*, *16*, 283–290. doi:10.3758/BF03203943
- Pomplun, M. (2006). Saccadic selectivity in complex visual search displays. *Vision Research*, *46*, 1886–1900. doi:10.1016/j.visres.2005.12.003
- Pomplun, M., & Sunkara, S. (2003). Pupil dilation as an indicator of cognitive workload in human-computer interaction. In D. Harris, V. Duffy, M. Smith, & C. Stephanidis (Eds.), *Proceedings of the 10th International Conference on Human-Computer Interaction: Vol. 3. Human-centred computing: Cognitive, social, and ergonomic aspects* (pp. 542–546). Mahwah, NJ: Erlbaum.
- Porter, G., Troscianko, T., & Gilchrist, I. D. (2007). Effort during visual search and counting: Insights from pupillometry. *Quarterly Journal of Experimental Psychology*, *60*, 211–229.
- Shore, D. I., & Klein, R. M. (2000). On the manifestations of memory in visual search. *Spatial Vision*, *14*, 59–75.
- Stanners, R. F., Coulter, M., Sweet, A. W., & Murphy, P. (1979). The pupillary response as an indicator of arousal and cognition. *Motivation and Emotion*, *3*, 319–340.
- Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*, 97–136. doi:10.1016/0010-0285(80)90005-5
- Vogel, E. K., & Machizawa, M. G. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature*, *428*, 748–751. doi:10.1038/nature02447
- Wolfe, J. M. (1994). Visual search in continuous, naturalistic stimuli. *Vision Research*, *34*, 1187–1195.