



# Tracking working memory maintenance with pupillometry

Nash Unsworth<sup>1</sup> · Matthew K. Robison<sup>1</sup>

Published online: 4 December 2017  
© The Psychonomic Society, Inc. 2017

## Abstract

Phasic pupillary responses were used to track the active maintenance of information in working memory (WM). In seven experiments participants performed various change detection tasks while their pupils were continuously recorded. Across the experiments phasic pupillary responses increased as the number of maintained items increased up to around 4–5 items consistent with behavioral estimates of capacity. Combining data across experiments demonstrated that phasic pupillary responses were related to behavioral estimates of capacity. Furthermore, phasic pupillary responses demonstrated WM load-dependent relations only when active maintenance was required. When instructed to passively stare at the items or to drop items from WM, the pupil remained near baseline levels. These phasic pupillary responses also tracked the time course of maintenance demonstrating sustained responses early in the delay period, but declined thereafter. Finally, phasic pupillary responses tracked selection processes at encoding (filtering and pre-cues), but did not suggest evidence for item removal following retro-cues. These results are consistent with the notion that maintaining items in WM requires the allocation of effortful attention and further suggest that phasic pupillary responses can be used to track the active maintenance of items in WM.

**Keywords** Working memory

Working memory is a system responsible for active maintenance and on-line manipulation of information. In our view, working memory (WM) consists of a subset of activated traces above threshold, some of which are highly active, strategies for maintaining activation of those traces, attention control processes to protect traces from interfering internal and external information, and controlled search processes to reactivate traces that could not be actively maintained. This view, like others, emphasizes the interaction of attention and memory in the service of complex cognition (Cowan, 2005; Engle & Kane, 2004; Unsworth & Engle, 2007). WM impairments have been demonstrated in psychiatric and neurological disorders associated with problems of disordered thought and forgetfulness such as schizophrenia, Alzheimer's, and Parkinson's (Gold et al., 2003; Lee et al., 2010; Parra et al., 2010). Even within a healthy population, WM ability reflects a core cognitive trait given its strong relations with higher-order cognition including intelligence and performance on

measures of scholastic aptitude (Engle et al., 1999; Unsworth et al., 2014).

An important aspect of WM is that it is thought to be capacity limited such that only roughly four items on average can be actively maintained (Cowan, 2001; Luck & Vogel, 2013). A great deal of recent research has demonstrated that there are important individual differences in the number of items that can be maintained in WM and variation in WM capacity is related to a number of cognitive abilities (e.g., Cowan et al., 2005; Vogel & Machizawa, 2004; Unsworth et al., 2014). For example, estimates of capacity are related to other WM measures and are related to broader measures of cognitive abilities such as fluid intelligence, attention control, and long-term memory abilities (Cowan et al., 2005; Fukuda et al., 2010; Shipstead et al., 2014; Unsworth et al., 2014). Recent research has also demonstrated that delay activity during visual WM tasks provides a neural correlate of WM capacity (e.g., Todd & Marois, 2004; Vogel & Machizawa, 2004). For example, Todd and Marois (2004) found that the delay signal in the intraparietal sulcus increased as set size increased, reaching asymptote around 3–4 items. Furthermore, Todd and Marois (2005) found that the delay activity was correlated with behavioral estimates of WM capacity. Similarly, Vogel and Machizawa (2004; Vogel, McCollough, & Machizawa, 2005) demonstrated that

---

✉ Nash Unsworth  
nashu@uoregon.edu

<sup>1</sup> Department of Psychology, University of Oregon,  
Eugene, OR 97403, USA

sustained activity over posterior parietal electrodes during the delay (the contralateral delay activity) of a visual WM task increased as set size increased and reached asymptote around 3–4 items. Importantly, the contralateral delay activity was strongly correlated with individual differences in behavioral estimates of WM capacity (see also Unsworth et al., 2015). These and other studies suggest that WM delay activity is a strong correlate of behavioral estimates of WM capacity.

Theoretically, the capacity limit arises because only four or so items can be individuated and maintained through the continued allocation of attention (Craik & Levy, 1976). Given this sharp capacity limit, it is critically important to encode and actively maintain only task relevant information to ensure fast and accurate responding. That is, given that only a few items can be maintained in WM at any given time, it is important to ensure that those items that are important for task performance are adequately encoded into WM and are actively maintained in WM during a delay. As such, the ability to actively maintain items in WM is critically dependent on the ability to allocate attention to items within WM. For example, research suggests that attention-based rehearsal processes may be needed to actively maintain information in visual WM via covert (or overt) shifts of attention to prioritized locations (Awh & Jonides, 2001; Awh, Vogel, & Oh, 2006). If attention is captured by distracting internal or external information, the representations will not be maintained and performance will suffer (see Allen et al., 2017 for recent evidence). Thus, it is important to consistently allocate attention to items in WM to prevent attentional capture from potent internal (e.g., mind-wandering) and external distraction. Although many theories suggest that attention is needed to actively maintain items in WM, the evidence supporting such a claim is mixed. Some studies have found dual-task costs when participants have to perform an attention demanding secondary task while maintaining items in WM (e.g., Allen et al., 2017; Morey & Bieler, 2013; Morey & Cowan, 2004, 2005), whereas others have found that participants can perform both tasks with little to no cost to WM performance (e.g., Belopolsky & Theeuwes, 2009; Fougne, 2009; Fougne & Marois, 2006; Hollingworth & Maxcey-Richard, 2013). Thus, it is not clear whether maintenance in WM is an active effortful process that requires the allocation of attention, or whether it is a passive effortless process that requires little attention (e.g., Fougne, 2009).

In the present study we suggest that pupil diameter can be used as a means to track effortful attention allocation and task engagement while performing WM tasks. Much prior research has shown that the pupil dilates in response to the cognitive demands of a task (Beatty, 1982). These effects reflect task-evoked phasic pupillary responses in which the pupil dilates relative to baseline levels due to increases in cognitive processing load. For example, Hess and Polt (1964) demonstrated that the pupils dilated as math problems became more difficult. A number of studies have demonstrated similar phasic

pupillary responses in a variety of tasks (Beatty & Lucero-Wagoner, 2000; see also Goldinger & Papesch, 2012; Laeng et al., 2012 for recent reviews). These and other results led Kahneman (1973) and Beatty (1982) to suggest that these phasic pupillary responses are reliable and valid psychophysiological markers of effortful attentional allocation (see also Alnaes et al., 2014; Daniels et al., 2012; Naber, Alvarez, & Nakayama, 2013). That is, phasic pupillary responses correspond to the intensive aspect of attention and provide an on-line indication of the utilization of capacity (Kahneman, 1973; Just & Carpenter, 1993).

Several studies have examined task-evoked pupillary responses during verbal WM tasks and have found that pupillary responses increase as the amount of information in WM increases (e.g., Heitz et al., 2008; Johnson, 1971; Johnson et al., 2014; Kahneman & Beatty, 1966; Peavler, 1974; Tsukahara et al., 2016). Recent research has also begun to examine pupillary responses in visual WM tasks where verbalization and rehearsal are less likely to influence estimates of WM capacity. For example, Unsworth and Robison (2015) had participants ( $N = 70$ ) perform a WM change detection task where the number of items to be maintained varied from 1 to 8 and the participants' pupils were measured continuously throughout the task. Specifically, participants were briefly presented with an array of colored squares followed by a delay period of 4,000 ms and then the test array. The participant's task was to indicate whether the circled item in the test array had changed its color from the memory array. Consistent with prior research, participants' WM capacity was estimated at close to four items (Cowan, 2001). Importantly, phasic pupillary responses (baseline corrected for each individual and each trial) increased as set size increased and then plateaued between four and five items consistent with the estimate of WM capacity. Furthermore, changes in the phasic pupillary response predicted individual estimates of capacity, suggesting that phasic pupillary responses provide a pupillary correlate of WM capacity similar to that found with contralateral delay activity (Vogel & Machizawa, 2004) and the fMRI signal in the intraparietal sulcus (Todd & Marois, 2004). The phasic responses also allowed us to track how participants effortfully allocated attention during the delay period. Specifically, when participants were required to maintain items below their capacity the pupil showed little dilation suggesting that individuals were allocating few attentional resources to maintain the items. When participants were asked to maintain a number of items at or above their capacity, however, the pupil ramped up and peaked early and then tended to maintain that level throughout the delay period suggesting that attention was being allocated in a more continuous manner to maintain the items in an active state. Similarly, Kursawe and Zimmer (2015) found that phasic pupillary responses increased as participants needed to maintain one, two, or four items and this did not seem to differ in terms of whether the

color, shape, or both were to-be-remembered. Additionally, in a luminance control condition, Kursawe and Zimmer presented participants with the same arrays, but participants were not required to remember them. In this task, phasic pupillary responses did not change as a function of the number of items in the array with very little dilation overall, suggesting that the phasic pupillary results were not simply due to the number of items being presented, but rather were due to the number of items that had to be actively maintained in WM. Collectively, these results suggest that it is possible to use phasic pupillary responses to track how participants actively maintain items in WM over short delays.

The goal of the current study was to use pupillary responses as an online measure of attentional allocation to better examine how attention is effortfully allocated to items in WM. Specifically, if phasic pupillary dilations provide an online measure of attentional allocation and capacity limits in WM reflect the number of items that can be maintained through the continued allocation of attention, then we should find that the pupil dilates up to around four items and then plateaus as more items are presented consistent with prior electrophysiological and fMRI research (Todd & Marois, 2004; Vogel & Machizawa, 2004) and prior pupillometry research (Kursawe & Zimmer, 2015; Unsworth & Robison, 2015). Furthermore, phasic pupillary responses should track the number of items being actively maintained in WM, but not necessarily track the number of items presented. That is, only when participants are required to hold onto and remember the items should the pupil track the number of items being maintained. In conditions where maintenance is not required, pupillary phasic responses should not change as a function of the number of items presented and should show very little dilation overall. Furthermore, in situations where only some (but not all) items need to be maintained, we should see that the pupil tracks the number of items to be maintained but not necessarily all presented items. Finally, assuming that attention needs to be continuously allocated to maintain items in WM, we can track the time course of the phasic dilations to determine if dilation maintains throughout the delay period. To examine these issues we conducted seven experiments where participants performed a number of WM change detection tasks while their pupils were measured continuously throughout the task. Using phasic pupillary dilations during the delay period of WM tasks as an index of effortful attention we should be able to track the allocation of attention to items in WM and provide a means of tracking the number of items held in WM on a moment-by-moment basis.

## Experiment 1

In Experiment 1, we examined whether the pupillary response during the delay period in prior research (e.g., Unsworth &

Robison, 2015) is due to active or passive maintenance. We have argued that maintaining items in WM is an active effortful process and this is the reason for the increase in pupil dilation as a function of set size. However, it is also possible that part of this response reflects a sensory load whereby the pupil is responding to changes in the sample array. Therefore, to examine whether the effects are due to active processes or due to passive maintenance of sensory responses participants performed the same visual arrays change detection task from Unsworth and Robison (2015). Prior to each trial participants were told whether the trial is an active trial or a passive trial. For active trials participants were told that they need to decide if the patterns were the same or different. On the passive trials they were told that they do not need to do anything, but just stare at the screen and simply press the space bar during the test screen. Thus, on active trials participants will theoretically have to maintain the array in WM until a response is required. On passive trials the array will not need to be maintained in WM. If the pupillary response is due to active effortful maintenance we should see the results in the active condition replicate prior research, but that in passive condition there should be no set size effect and no increase in pupil dilation (e.g., Kursawe & Zimmer, 2015). Furthermore, comparing set sizes in the two conditions it is possible that for small set sizes there will be no differences across conditions suggesting that when maintaining only one or two items this can be done passively. However, as the number of items increase up to one's capacity maintenance becomes more effortful leading to differences in the pupillary response for larger set sizes.

## Method

### Participants

Participants were 39 undergraduate students recruited from the subject pool at the University of Oregon. Participants were between the ages of 18 and 35 years and received course credit for their participation. Data from two participants were excluded from analyses because of data collection problems with the eye-tracker leaving a final sample of 37 participants.

### Procedure

Participants were tested individually in a dark room. Pupil diameter was continuously recorded binocularly at 120 Hz using a Tobii T120 eye-tracker. Participants were seated 60 cm from the monitor. After providing informed consent and after calibrating the eye-tracker, participants performed a change detection task. In this task participants were first presented with a black fixation cross in the middle of the screen on a gray background for 2,000 ms. Next participants were presented with arrays of 1–8 colored squares ( $0.65^\circ \times 0.65^\circ$ )

for 250 ms. The arrays were arranged randomly on a neutral gray background with each color randomly selected from one of seven easily discriminable colors (red, blue, violet, green, yellow, black, or white). The items in the arrays were separated by at least  $2^\circ$  of visual angle measured from the centers of the square. The presentation of the arrays was followed by a delay period of 4,000 ms and finally the test array reappeared with one of the items circled. Participants responded as to whether or not the circled item had changed color. Half of the trials were change trials. Twenty trials of each array size were randomly presented for a total of 160 trials. Prior to each trial participants were told whether the trial was an active trial or a passive trial. Specifically, the word Active or the word Passive appeared onscreen for 1,500 ms. Half of the trials were active and half were passive, with trial type randomly intermixed. For active trials participants were told that they need to decide if the patterns are the same or different. On the passive trials they were told that they do not need to do anything, but just stare at the screen and when the test array appears press the space bar.

### Pupil data analysis

Data from each participant's left eye was used (left and right eye pupil diameter were highly correlated  $r = .95$ ). All trials (correct and error trials) were examined. Missing data points due to blinks, off-screen fixations, and/or eye-tracker malfunction were removed (roughly 16.5 % of the overall data with roughly equal amounts of data loss across conditions and set sizes). The relevant period of missing data was not included in the averaging. Phasic responses were baseline corrected by subtracting out the last 500 ms of the baseline pupil diameter during fixation on a trial-by-trial basis for each participant. During the delay period the pupil data were averaged into a series of 200-ms time windows for each trial and each 200-ms window was baseline corrected.

## Results and discussion

### Accuracy and K estimates

First we examined accuracy as a function of set size. As shown in Table 1, accuracy was high when four or fewer items were present, but steadily decreased with larger set sizes,  $F(7, 252) = 20.25$ ,  $MSE = .02$ ,  $p < .001$ , partial  $\eta^2 = .36$ . K was estimated using Cowan's (2001) formula for each set size and each individual. The values for set sizes 4–8 were then averaged to get an estimate of capacity. Across all individuals the K estimate was 3.4 ( $SD = 1.46$ ), which was significantly different from zero,  $t(36) = 13.54$ ,  $p < .001$ .

**Table 1** Proportion correct as a function of set size and experiment

Experiment	Set Size							
	1	2	3	4	5	6	7	8
E1 active condition	.94	.91	.89	.89	.85	.79	.74	.69
E2 hold condition	.94	.93	.91	.90	.85	.81	.77	.68
E3a 4,000-ms delay	.95	.95	.92	.91	.87	.83	.75	.75
E3a 8,000-ms delay	.97	.93	.90	.90	.82	.81	.74	.67
E3b 4,000-ms delay	.97	.92	.95	.93	.87	.84	.78	.71
E3b 8,000-ms delay	.96	.98	.98	.95	.91	.87	.80	.81

### Pupillary responses

Next we examined phasic pupillary responses for the active and passive conditions. Phasic pupillary responses were submitted to a 2 (Active vs. Passive)  $\times$  8 (Set Size)  $\times$  20 (200 ms Bin) repeated measures ANOVA. The ANOVA yielded a main effect of type,  $F(1, 36) = 16.49$ ,  $MSE = .27$ ,  $p < .001$ , partial  $\eta^2 = .31$ , with greater dilation for the active than the passive condition. There was a main effect of set size,  $F(7, 252) = 3.53$ ,  $MSE = .06$ ,  $p = .001$ , partial  $\eta^2 = .09$ , with larger dilations for larger set sizes. There was also a main effect of bin,  $F(19, 684) = 6.58$ ,  $MSE = .01$ ,  $p < .001$ , partial  $\eta^2 = .16$ . The type  $\times$  set size interaction approached conventional levels of significance,  $F(7, 252) = 1.91$ ,  $MSE = .06$ ,  $p = .069$ , partial  $\eta^2 = .05$ . The type  $\times$  bin interaction was significant,  $F(19, 684) = 4.54$ ,  $MSE = .004$ ,  $p < .001$ , partial  $\eta^2 = .11$ , as was the set size  $\times$  bin interaction,  $F(133, 4788) = 2.35$ ,  $MSE = .003$ ,  $p < .001$ , partial  $\eta^2 = .06$ . Importantly, the type  $\times$  set size  $\times$  bin interaction was significant,  $F(133, 4788) = 1.96$ ,  $MSE = .005$ ,  $p < .001$ , partial  $\eta^2 = .05$ . Decomposing this interaction suggested that in the Active condition there was a significant main effect of set size,  $F(7, 252) = 4.42$ ,  $MSE = .06$ ,  $p < .001$ , partial  $\eta^2 = .11$ , and a set size  $\times$  bin interaction,  $F(133, 4788) = 2.64$ ,  $MSE = .003$ ,  $p < .001$ , partial  $\eta^2 = .07$ . As seen in Fig. 1a, in the Active condition pupil diameter increased as set size increased and then plateaued at four items consistent with prior research (Unsworth & Robison, 2015). That is, there was no effect of set size for set sizes 4–8,  $F(4, 144) = .72$ ,  $MSE = .055$ ,  $p = .58$ , partial  $\eta^2 = .02$ . Furthermore, as shown in Fig. 1b, there were clear differences in the phasic pupillary responses for small versus large set sizes with the small set sizes demonstrating little dilation during the delay period whereas the larger set sizes demonstrated larger dilations that maintained throughout the delay period replicating prior research (Unsworth & Robison, 2015). Specifically, examining set sizes 1 and 2 compared to set sizes 7 and 8 in the Active condition suggested an overall main  $F(1, 36) = 13.95$ ,  $MSE = .076$ ,  $p = .001$ , partial  $\eta^2 = .28$ , with the larger set sizes demonstrating larger overall changes in pupil diameter ( $M = .042$ ,  $SE = .01$ ) than the smaller set sizes ( $M = .004$ ,  $SE = .01$ ). There was also an interaction,  $F(19, 684) = 8.93$ ,  $MSE =$

.003,  $p < .001$ , partial  $\eta^2 = .20$ , indicating larger phasic responses for the larger set sizes compared to the small set sizes. For the Passive condition, however, there was no main effect of set size,  $F(7, 252) = 0.90$ ,  $MSE = .06$ ,  $p = .50$ , partial  $\eta^2 = .02$ . As shown in Fig. 1a, there was no set size effect for the Passive condition. For the Passive condition there was a set size  $\times$  bin interaction,  $F(133, 4788) = 1.61$ ,  $MSE = .002$ ,  $p < .001$ , partial  $\eta^2 = .04$ . As shown in Fig. 1c, all of the set sizes demonstrated similar phasic responses that were small and resembled the phasic responses for small set sizes in the Active condition. The interaction seemed to be mainly driven by the larger constriction response for Set Size 1 compared to the other set sizes. No other effects were significant.

Overall, the results from the Active condition replicate prior research demonstrating that phasic pupillary responses track the number of items being maintained in WM. Results from the Passive condition replicated prior research suggesting that when the same arrays of items are presented, but there is no requirement for maintenance, there is little phasic dilation and little to no differences across set size (Kursawe & Zimmer, 2015). Furthermore, for small set sizes similar phasic responses were seen in the Active and Passive condition suggesting that when the number of items that need to be maintained is small (only one or two items), there is little effort invested in maintaining these items. For example, examining set sizes 1 and 2 in the Active and Passive conditions suggested no condition  $\times$  bin interaction,  $F(19, 684) = 0.50$ ,  $MSE = .005$ ,  $p = .97$ , partial  $\eta^2 = .01$ , indicating overall similar phasic responses. Although it should be noted that there was a main effect of condition,  $F(1, 36) = 4.20$ ,  $MSE = .007$ ,  $p = .048$ , partial  $\eta^2 = .10$ , with the Active condition demonstrating larger overall changes in pupil diameter ( $M = .004$ ,  $SE = .009$ ) than the Passive condition ( $M = -.016$ ,  $SE = .010$ ). Collectively, these results are consistent with the notion that phasic pupillary responses track the number of items that are actively maintained in WM during the delay period and do not simply provide an index of sensory load carrying over from the presentation of the arrays.

## Experiment 2

Experiment 2 examined whether it is possible to start out actively trying to maintain items in WM, but then dropping those items if requested to do so. Furthermore, we examined whether phasic pupillary responses would track this change in maintenance. Participants performed the same visual arrays task as before. On half of the trials halfway through the delay period participants heard “hold” indicating that they needed to maintain the array throughout the whole delay. On the other half of trials halfway through the delay period participants heard “drop” indicating that they do not need to remember the array. For the Hold trials we should see that the pupil dilates early in

the delay period and remains dilated throughout the delay consistent with Experiment 1 and prior research. On the Drop trials, however, the pupil should ramp up early and maintain for the first part of the delay, but following the drop signal the pupil should decrease back to baseline levels indicating that the items were no longer being actively maintained in WM.

## Method

### Participants

Participants were 48 undergraduate students recruited from the subject pool at the University of Oregon. Participants were between the ages of 18 and 35 years and received course credit for their participation. Data from 11 participants were excluded from analyses because of data collection problems with the eye-tracker and data from three participants were excluded due to having accuracy below 50 % leaving a final sample of 34 participants.

### Procedure

This was the same as for Experiment 1 with the following exceptions. On half of the trials halfway through the delay period (2,500 ms) participants heard “hold” indicating to maintain the array throughout the whole delay. On the other half of trials halfway through the delay period (2,500 ms) participants heard “drop” indicating that they could drop the array. In both trial types the overall delay period was 5,200 ms including the time taken to hear “hold” or “drop.” For Hold trials participants were told that they needed to decide if the patterns are the same or different. On the Drop trials they were told that they do not need to do anything, but just stare at the screen and when the test array appears press the space bar. Half of the trials were change trials. Sixteen trials of each array size were randomly presented for a total of 128 trials.

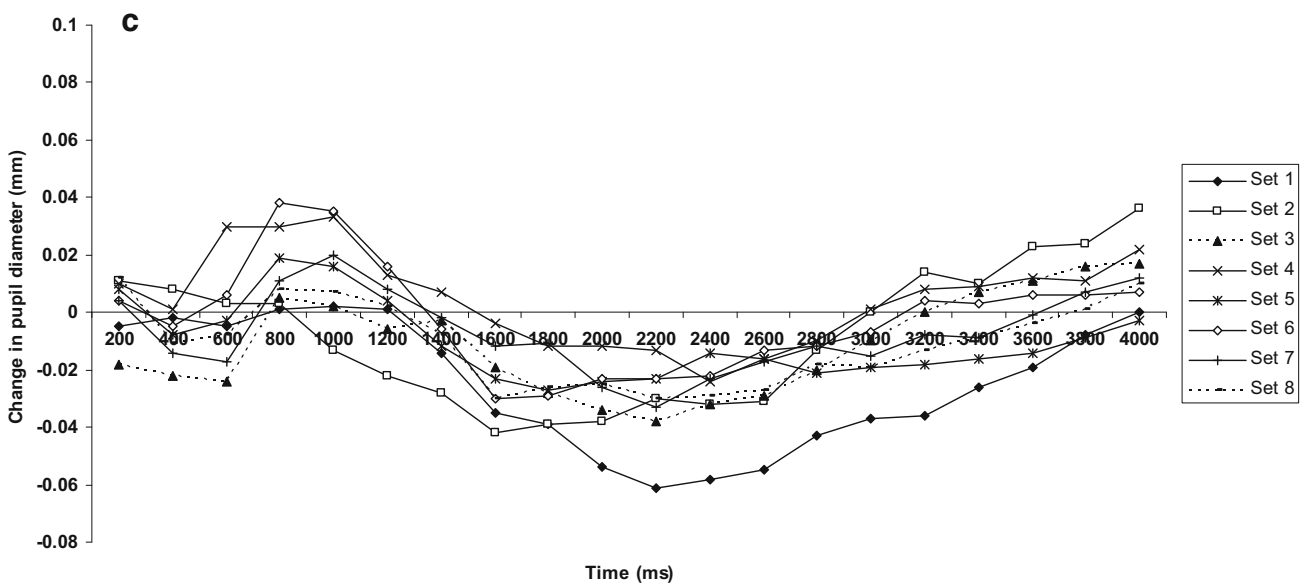
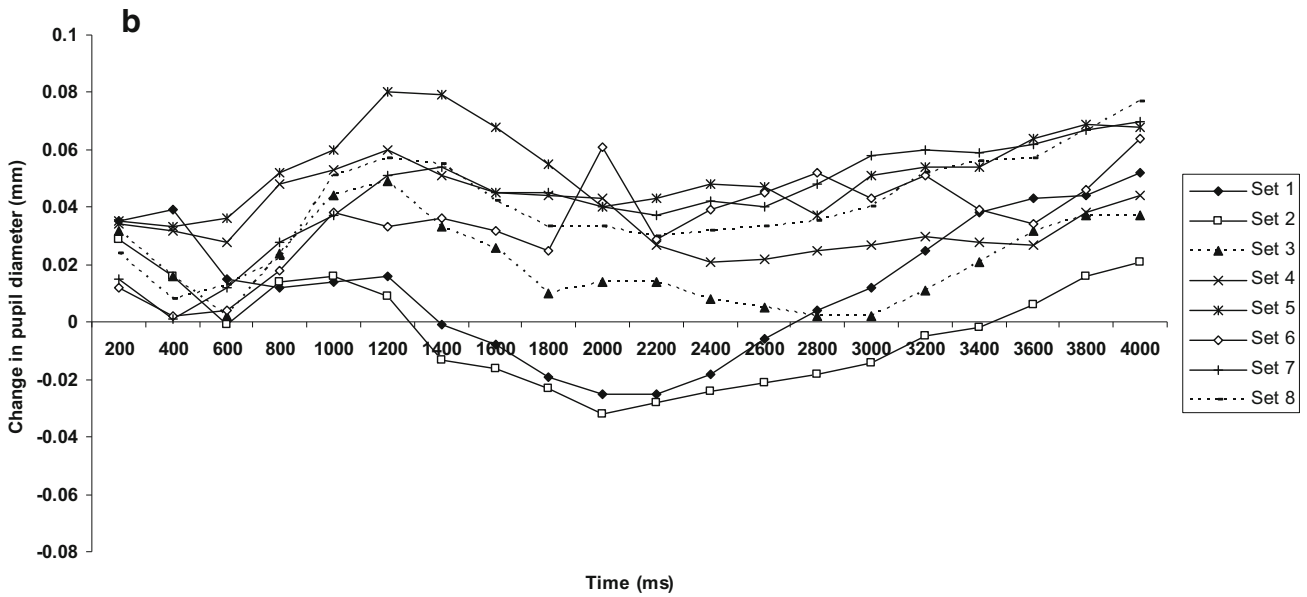
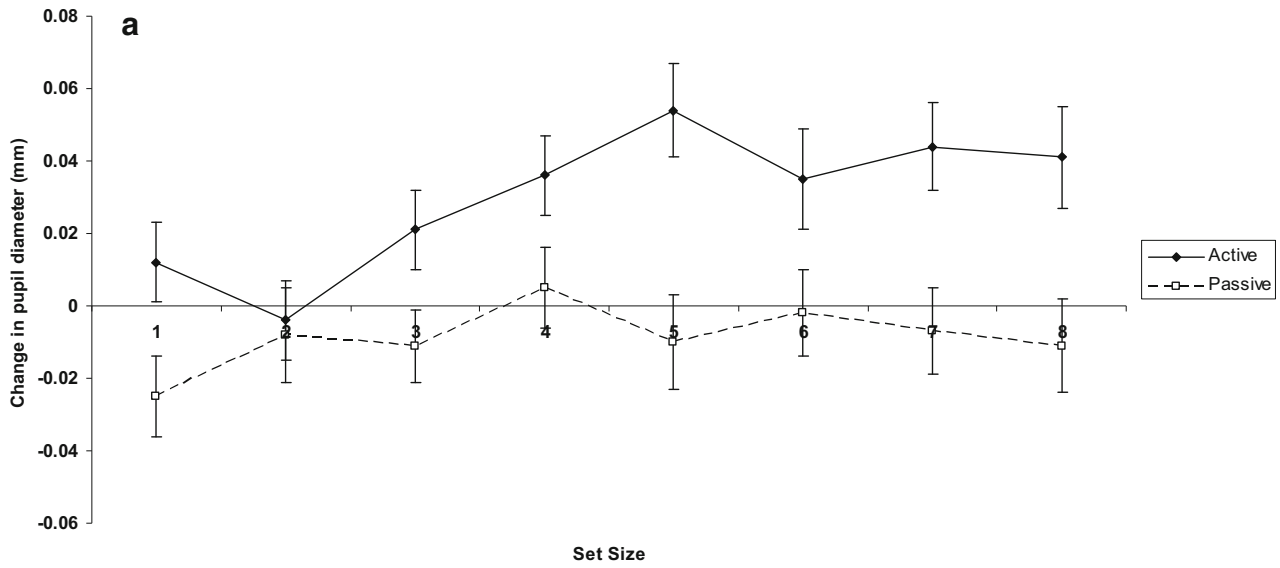
### Pupil data analysis

This was the same as for Experiment 1. Missing data points due to blinks, off-screen fixations, and/or eye-tracker malfunction were removed (roughly 20.3 % of the overall data with roughly equal amounts of data loss across conditions and set sizes).

## Results and discussion

### Accuracy and K estimates

As shown in Table 1, accuracy was high when four or fewer items were present, but steadily decreased with larger set sizes,  $F(7, 231) = 16.59$ ,  $MSE = .02$ ,  $p < .001$ , partial  $\eta^2 = .33$ . Across



**Fig. 1** Experiment 1. (a) Change in pupil diameter during the delay as a function of set size and Active vs. Passive condition. Error bars reflect one standard error of the mean. (b) Change in pupil diameter as a function of set size and time point during the delay in the Active condition. (c) Change in pupil diameter as a function of set size and time point during the delay in the Passive condition

all individuals the K estimate was 3.26 ( $SD = 1.59$ ), which was significantly different from zero,  $t(33) = 11.98, p < .001$ .

## Pupillary responses

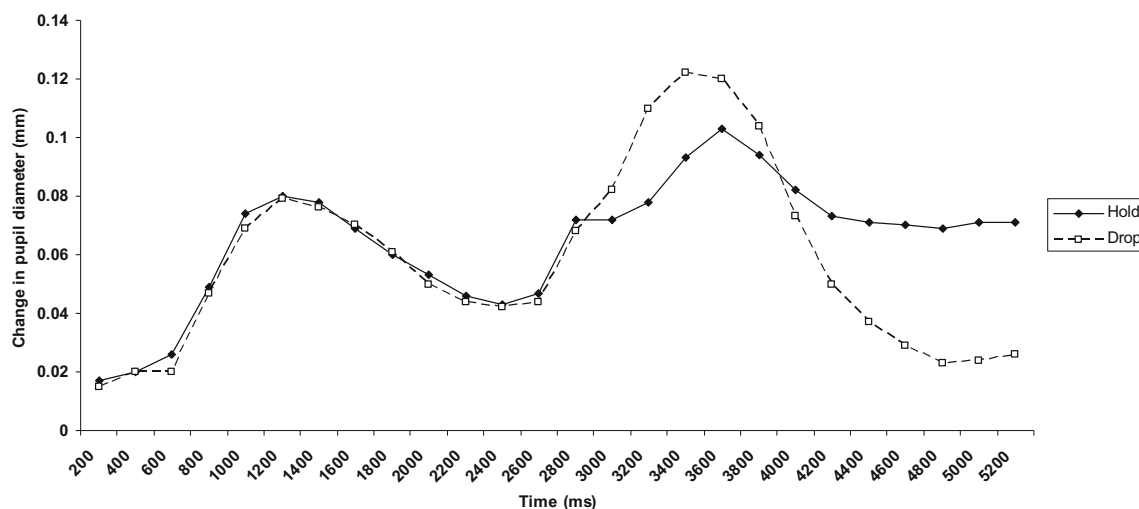
Next we examined phasic pupillary responses for the Hold and Drop conditions. Phasic pupillary responses were submitted to a 2 (Hold vs. Drop)  $\times$  8 (Set Size)  $\times$  26 (200 ms Bin) repeated measures ANOVA. There was a main effect of set size,  $F(7, 231) = 7.52, MSE = .18, p < .001, \text{partial } \eta^2 = .19$ . Consistent with Experiment 1 and prior research, pupil dilation increased up to set size four and then plateaued. There was also a main effect of bin,  $F(25, 825) = 14.42, MSE = .02, p < .001, \text{partial } \eta^2 = .30$ . There was also a set size  $\times$  bin interaction,  $F(175, 5775) = 4.84, MSE = .005, p < .001, \text{partial } \eta^2 = .13$ , consistent with Experiment 1 and prior research. Importantly, the type  $\times$  bin interaction was significant,  $F(25, 825) = 7.59, MSE = .008, p < .001, \text{partial } \eta^2 = .19$ . No other effects were significant. As shown in Fig. 2, the Hold and Drop trials were similar until the auditory signal to Hold or Drop the array. For Hold trials phasic pupil response maintained after the signal, but for Drop trials the phasic pupil response dropped back to baseline levels following the signal. Examining the last time bin suggested that Hold and Drop trials differed significantly from each other,  $t(33) = 3.31, p = .002, d = .57$ . Furthermore, Hold trials differed significantly from baseline,  $t(33) = 5.22, p < .001$ , but Drop trials did not,  $t(33) = .89, p = .379$ . Additionally, note that the response to the Drop signal is initially larger than the response to the hold

signal,  $t(33) = 2.19, p = .035, d = .39$ . This could reflect an active process of removing the items, which presumably requires effort, or it could reflect differences in the pupil response to the slightly different words. Additional research is required to better examine potential pupillary correlates of active removal processes.

Similar to Experiment 1, the results from Experiment 2 suggest that phasic pupillary responses track the number of items being actively maintained in WM. When told to hold onto items, phasic pupillary responses maintained throughout the delay period. However, when told to drop items, the phasic pupillary response dropped back to baseline levels suggesting that participants were no longer using attentional effort to maintain the items in WM. These results are similar to the unloading function described by Kahneman and Beatty (1966) and what happens following a directed forgetting cue (Johnson, 1971), suggesting that the load on WM is being reduced. Consistent with Experiment 1 these results suggest that phasic pupillary responses not only track the number of items being maintained in WM, they also track whether and when participants are actively maintaining items in WM.

## Experiment 3a

The prior experiments suggested that the pupil sustained dilation throughout the delay period as if participants were actively maintaining the items throughout the entire delay. However, it is not known how long participants can actively maintain the information. Prior pupillary research with sustained tasks like multi-object tracking (which is closely related to WM; e.g., Drew & Vogel, 2008) has shown that the pupil response can be maintained for several seconds as long as participants are still actively using the information (Alnaes et al., 2014; Wright, Boot, & Morgan, 2013). However, in such tasks attention is



**Fig. 2** Change in pupil diameter during the delay for Hold vs. Drop trials for Experiment 2

directed to perceptual stimuli that are still present. For WM, attention is directed inward to memory representations that are no longer present in the environment. Experiment 3 examined whether participants can continuously allocate attention to items in WM over longer delays than previously used. Participants performed the visual arrays task as before. On half of the trials participants maintained items over a 4,000-ms delay and on the other half of trials participants were required to maintain the items over an 8,000-ms delay. The 4,000-ms delay condition should replicate the prior experiments. For the 8,000-ms condition it is possible that participants can continue to maintain the items leading to a sustained pupillary response or it is possible that during some time point the pupil will return to baseline indicating that participants could no longer effortfully maintain the items in WM.

## Method

### Participants

Participants were 24 undergraduate students recruited from the subject pool at the University of Oregon. Participants were between the ages of 18 and 35 years and received course credit for their participation. Data from one participant were excluded from analyses because of data collection problems with the eye-tracker leaving a final sample of 23 participants.

### Procedure

This was the same as for Experiment 1 with the following exceptions. On half of the trials the delay period was 4,000 ms and on the other half of trials the delay period was 8,000 ms. The two delay intervals were randomly intermixed within participants. Half of the trials were change trials. Twenty trials of each array size were randomly presented for a total of 160 trials.

### Pupil data analysis

This was the same as for Experiment 1. Missing data points due to blinks, off-screen fixations, and/or eye-tracker malfunction were removed (roughly 17.3 % of the overall data with roughly equal amounts of data loss across conditions and set sizes).

## Results and discussion

### Accuracy and K estimates

As shown in Table 1, accuracy was high when four or fewer items were present, but steadily decreased with larger set sizes,  $F(7, 154) = 29.86$ ,  $MSE = .01$ ,  $p < .001$ , partial  $\eta^2 = .58$ , and this did not change as a function of delay interval,  $F(7, 154) =$

$.751$ ,  $MSE = .01$ ,  $p = .629$ , partial  $\eta^2 = .03$ . There were no differences between the two conditions in terms of overall accuracy,  $F(1, 22) = 2.93$ ,  $MSE = .02$ ,  $p = .101$ , partial  $\eta^2 = .12$ . K estimates were similar for the 4,000-ms 3.72 ( $SD = 1.63$ ) and 8,000-ms conditions, 3.23 ( $SD = 1.65$ ),  $t(33) = 1.68$ ,  $p = .106$ ,  $d = .35$ .

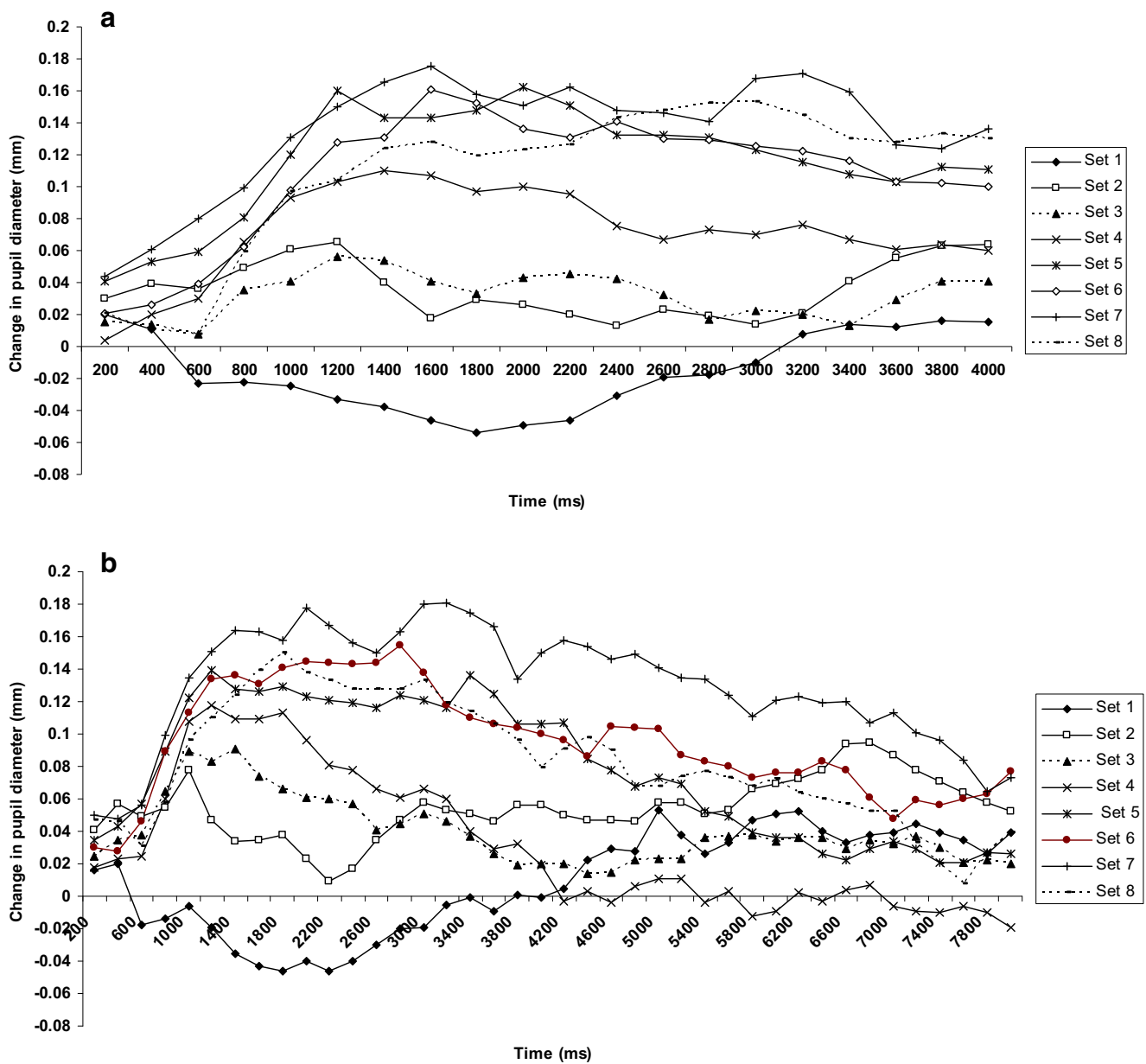
### Pupillary responses

Next we examined phasic pupillary responses for the first 4,000 ms of the delay period for both the 4,000-ms and 8,000-ms delay conditions to see if differences arose as a function of delay condition. Consistent with the prior experiments there was a main effect of set size,  $F(7, 154) = 13.00$ ,  $MSE = .19$ ,  $p < .001$ , partial  $\eta^2 = .37$ , in which the pupil dilated up to set size four and five and then plateaued. There was also a main effect of bin,  $F(19, 418) = 6.58$ ,  $MSE = .02$ ,  $p < .001$ , partial  $\eta^2 = .23$ . There was also a set size  $\times$  bin interaction,  $F(133, 2926) = 5.90$ ,  $MSE = .005$ ,  $p < .001$ , partial  $\eta^2 = .21$ , consistent with the prior experiments and prior research. For example, shown in Fig. 3a are the phasic pupillary responses for each set size across the delay period in the 4,000-ms delay condition. As can be seen, the phasic pupillary responses increased across set size stabilizing around Set Size 5 and these dilations maintained throughout the delay period consistent with the prior experiments and prior research. The only effect involving delay interval to approach conventional levels of significance was the delay  $\times$  bin interaction,  $F(19, 418) = 1.49$ ,  $MSE = .003$ ,  $p = .085$ , partial  $\eta^2 = .06$ .

Next, we specifically examined the phasic pupillary responses across the entire 8,000 ms delay for the 8,000-ms condition. There was a main effect of set size,  $F(7, 154) = 3.74$ ,  $MSE = .37$ ,  $p = .001$ , partial  $\eta^2 = .15$ . There was a main effect of bin,  $F(39, 858) = 4.49$ ,  $MSE = .02$ ,  $p < .001$ , partial  $\eta^2 = .17$ . There was also a set size  $\times$  bin interaction,  $F(273, 6006) = 4.12$ ,  $MSE = .005$ ,  $p < .001$ , partial  $\eta^2 = .16$ . As shown in Fig. 3b, most of the set sizes (except Set Size 1) showed sustained dilation for the first 4,000 ms or so. However, following that all of the set sizes demonstrated significant reductions in the phasic responses such that by the end of the delay period all of the phasic responses were back near baseline. In fact, in all set sizes the last time bin was not significantly different from baseline, all  $t$ 's  $< 1.5$ , all  $p$ 's  $> .15$ .

These results suggest that unlike for multi-object tracking (Alnaes et al., 2014; Wright, Boot, & Morgan, 2013), phasic pupillary responses were not sustained over longer delay periods when maintaining items in WM. Rather after 4,000 ms or so the phasic pupillary response begins to decline, reaching baseline levels near the end of the delay period. This suggests that participants are no longer using attentional effort to actively maintain the items in WM. This could be because the items have been transferred to



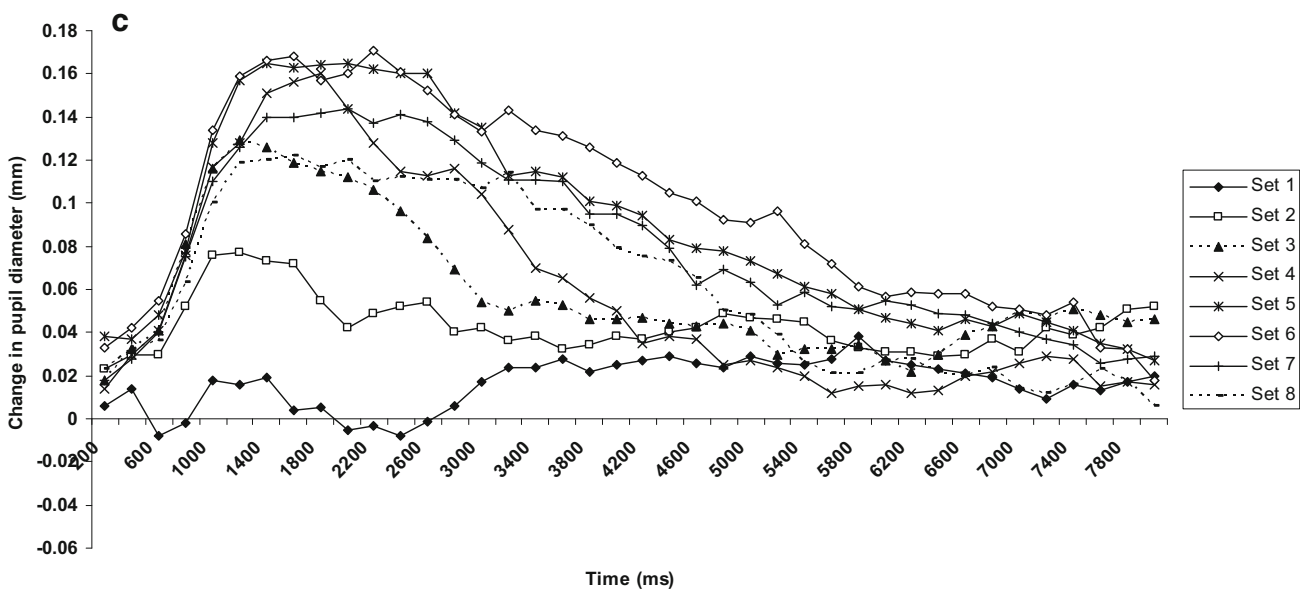
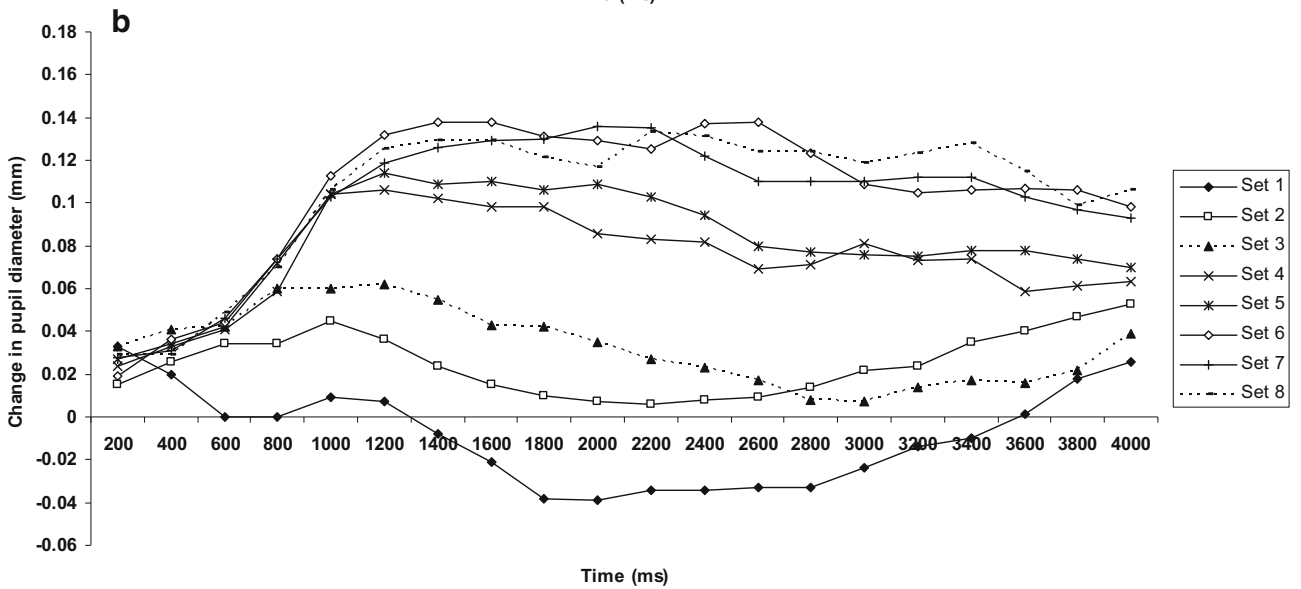
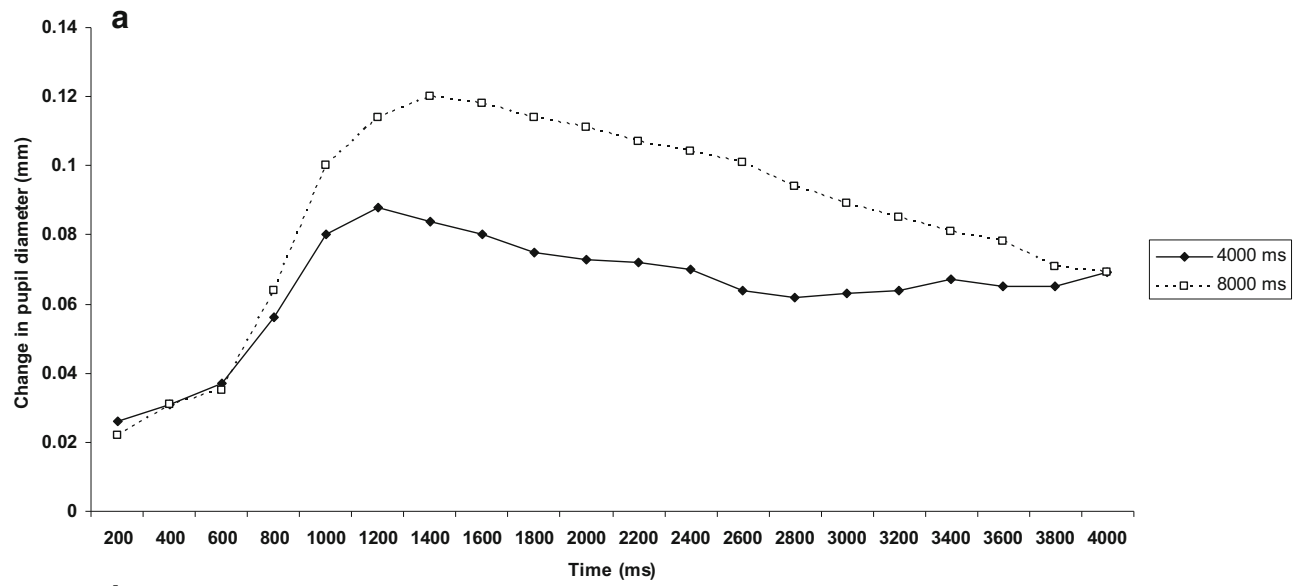


**Fig. 3** Experiment 3a. (a) Change in pupil diameter as a function of set size and time point during the delay in the 4,000-ms delay condition. (b) Change in pupil diameter as a function of set size and time point during the delay in the 8,000-ms delay condition

long-term memory, the items are decaying in WM are and are no longer being maintained (although the accuracy results suggested no differences between conditions), participants switch strategies in how they are maintaining the items, or the items are being held in a more passive fashion. The current results cannot speak to which of these possibilities are correct, but they do indicate that participants are no longer engaging in active effortful processes to maintain the items in WM. Whether this effect represents a limitation of the system or a strategic decision on the part of the participant is unclear and suggests future research is needed to better examine this effect.

### Experiment 3b

Experiment 3a suggested that phasic pupillary responses were not sustained over the entire delay in the 8,000-ms condition, but rather started to decline after approximately 4,000 ms. One problem with these results is that the 4,000-ms and 8,000-ms delay conditions were intermixed, and thus participants did not know what delay interval they would be receiving on each trial. This could have resulted in participants treating each trial as a 4,000-ms delay trial. In Experiment 3b the delay conditions were blocked to determine if knowing the delay duration influences the results in



◀ **Fig. 4** Experiment 3b. (a) Change in pupil diameter as a function of delay condition and time point during the delay. (b) Change in pupil diameter as a function of set size and time point during the delay in the 4,000-ms delay condition. (c) Change in pupil diameter as a function of set size and time point during the delay in the 8,000-ms delay condition

terms of whether the phasic pupillary response can be sustained in the longer 8,000-ms condition.<sup>1</sup>

## Method

### Participants

Participants were 30 undergraduate students recruited from the subject pool at the University of Oregon. Participants were between the ages of 18 and 35 years and received course credit for their participation.

### Procedure

This was the same as for Experiment 3a except that delay condition was blocked and the order of delay conditions was counterbalanced across participants. Sixteen trials of each array size were randomly presented for a total of 128 trials.

### Pupil data analysis

This was the same as for Experiment 1. Missing data points due to blinks, off-screen fixations, and/or eye-tracker malfunction were removed (roughly 9.3 % of the overall data with roughly equal amounts of data loss across conditions and set sizes).

## Results and discussion

### Accuracy and K estimates

As shown in Table 1, there was an effect of set size,  $F(7, 203) = 31.42$ ,  $MSE = .01$ ,  $p < .001$ , partial  $\eta^2 = .52$ , and this did not change as a function of delay interval,  $F(7, 203) = 1.49$ ,  $MSE = .009$ ,  $p = .173$ , partial  $\eta^2 = .05$ . There was, however, a main effect of delay,  $F(1, 29) = 8.39$ ,  $MSE = .02$ ,  $p = .007$ , partial  $\eta^2 = .22$ , suggesting that accuracy in the 8,000-ms condition ( $M = .91$ ,  $SE = .01$ ) was higher than in the 4,000-ms condition ( $M = .87$ ,  $SE = .02$ ). Similarly, K estimates were higher for the 8,000-ms condition ( $M = 4.35$ ,  $SE = .24$ ) than for the 4,000-ms condition ( $M = 3.62$ ,  $SE = .24$ ),  $t(33) = 2.57$ ,  $p = .016$ ,  $d = .47$ .

<sup>1</sup> We thank Keisuke Fukuda for suggesting this experiment.

## Pupillary responses

Next we examined phasic pupillary responses for the first 4,000 ms of the delay period for both the 4,000-ms and 8,000-ms delay conditions to see if differences arose as a function of the delay condition. There was a main effect of set size,  $F(7, 203) = 14.34$ ,  $MSE = .14$ ,  $p < .001$ , partial  $\eta^2 = .33$ . There was a main effect of bin,  $F(19, 551) = 19.88$ ,  $MSE = .01$ ,  $p < .001$ , partial  $\eta^2 = .41$ . There was a main effect of delay,  $F(1, 29) = 4.47$ ,  $MSE = .23$ ,  $p = .043$ , partial  $\eta^2 = .13$ , in which phasic responses were larger for the 8,000-ms condition ( $M = .09$ ,  $SE = .01$ ) than for the 4,000-ms condition ( $M = .07$ ,  $SE = .01$ ). There was also a set size  $\times$  bin interaction,  $F(133, 3857) = 8.68$ ,  $MSE = .003$ ,  $p < .001$ , partial  $\eta^2 = .23$ . There was also a delay  $\times$  bin interaction,  $F(19, 551) = 6.28$ ,  $MSE = .005$ ,  $p < .001$ , partial  $\eta^2 = .18$ . As shown in Fig. 4a, there was a bigger ramp up early on in the phasic pupil response for the 8,000-ms condition compared to the 4,000-ms condition. Finally, there was a significant delay  $\times$  set size  $\times$  bin interaction,  $F(133, 3857) = 1.36$ ,  $MSE = .002$ ,  $p = .005$ , partial  $\eta^2 = .05$ . As shown in Figs. 4b and c, the larger ramp up in phasic dilation in the 8,000-ms condition compared to the 4,000-ms condition was primarily due to a larger increase in phasic dilation for smaller set sizes. Specifically, for set sizes 1–5 there was a significant delay  $\times$  bin interaction suggesting a larger ramp up in the phasic response for the 8,000-ms condition than the 4,000-ms condition, all  $F$ 's  $> 2.5$ , all  $p$ 's  $< .001$ . However, for set sizes 6–8 there were no delay  $\times$  bin interactions suggesting similar phasic responses for the 8,000- and 4,000-ms conditions, all  $F$ 's  $< .68$ , all  $p$ 's  $> .64$ .

Next, we specifically examined the phasic pupillary responses across the entire 8,000-ms delay for the 8,000-ms condition. There was a main effect of set size,  $F(7, 203) = 3.42$ ,  $MSE = .24$ ,  $p = .002$ , partial  $\eta^2 = .11$ . There was a main effect of bin,  $F(39, 1131) = 22.55$ ,  $MSE = .01$ ,  $p < .001$ , partial  $\eta^2 = .44$ . There was also a set size  $\times$  bin interaction,  $F(273, 7917) = 4.29$ ,  $MSE = .004$ ,  $p < .001$ , partial  $\eta^2 = .13$ . As shown in Fig. 4c, and consistent with Experiment 3a, most of the set sizes (except Set Size 1) showed increased dilation early, followed by significant reductions in the phasic responses such that by the end of the delay period all of the phasic responses were back near baseline. Consistent with Experiment 3a in all set sizes the last time bin was not significantly different from baseline, all  $t$ 's  $< 1.3$ , all  $p$ 's  $> .23$ .

Consistent with Experiment 3a, the results of Experiment 3b suggest that the phasic pupillary responses were not sustained over longer delay periods when maintaining items in WM. In the blocked 4,000-ms condition, the phasic response sustained during the delay period consistent with prior results. In the blocked 8,000-ms condition, the initial phasic response was greater than for the 4,000-ms condition (especially for smaller set sizes). This suggests that knowing the delay interval allows participants to allocate more attentional

effort early in the delay period, perhaps in an attempt to overcome any loss over the longer delay. However, in the 8,000-ms blocked condition, soon after peaking the phasic pupillary response began to decline, reaching baseline levels near the end of the delay period. This suggests that participants are not sustaining attentional effort to actively maintain the items throughout the entire long delay interval.

### Experiment 4

Experiment 4 examined the extent to which participants can select which items are encoded into WM and whether the pupillary response reflects the number of relevant

items being maintained or the overall number of items presented. Participants performed a visual arrays task in which red and blue rectangles of different orientations were presented. Participants were told to remember the orientations of the red rectangles and ignore the blue rectangles. Thus, participants had to select the relevant category and filter out the irrelevant items. Prior research with this task has suggested that participants can indeed select the target items (Vogel et al., 2005). The question here is whether the phasic pupillary response will reflect the number of target representations to be maintained. Participants will be presented with two targets alone, four targets alone, two targets and two distractors, or four targets and two distractors. If participants can effectively

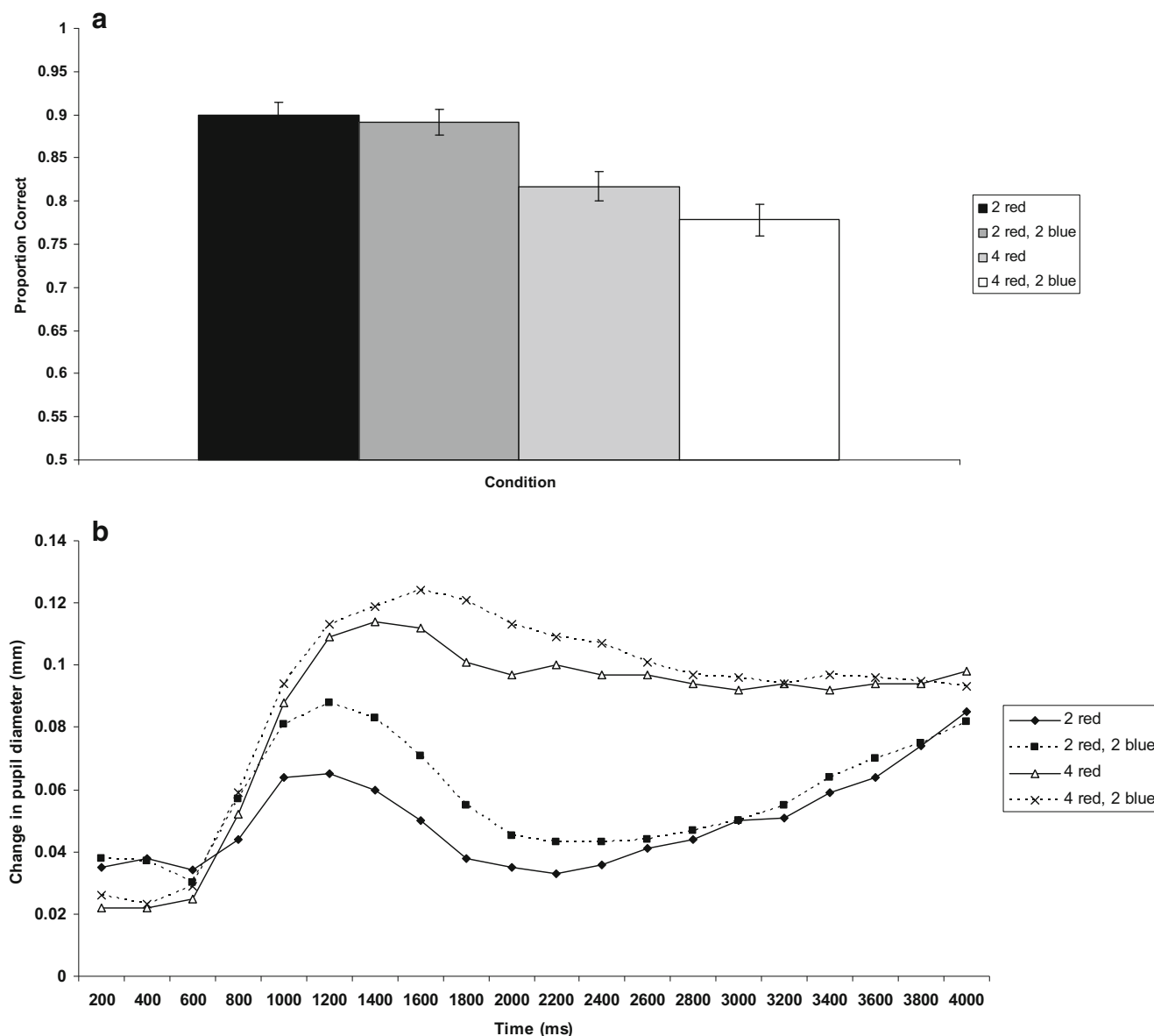


Fig. 5 Experiment 4. (a) Proportion correct as a function of each condition. Error bars reflect one standard error of the mean. (b) Change in pupil diameter as a function of target set size, presence of distractors, and time point during the delay

filter out the distractors we should see that the pupillary response is the same for conditions with targets alone and with targets and distractors. If participants cannot filter out the distractor, then the pupillary response should reflect the total number of items presented rather than the number of targets presented.

## Method

### Participants

Participants were 49 undergraduate students recruited from the subject pool at the University of Oregon. Participants were between the ages of 18 and 35 years and received course credit for their participation. Data from nine participants were excluded from analyses because of data collection problems with the eye-tracker and data from three participants were excluded due to having accuracy of approximately 50 % leaving a final sample of 37 participants.

### Procedure

Participants performed a change detection task modeled after Vogel et al. (2005). In this task participants were first presented with a black fixation cross in the middle of the screen on a gray background for 2,000 ms. Next participants were presented with arrays of two, four, or six red and/or blue rectangles for 250 ms. Arrays consisted of two red rectangles alone, two red rectangles and two blue rectangles, four red rectangles alone, or four red rectangles and two blue rectangles. Participants were instructed to remember the orientations of the red rectangles and ignore the blue rectangles. Items were presented within a gray  $19.1^\circ \times 14.3^\circ$  field. Items were separated from one another by at least  $2^\circ$  and were all at least  $2^\circ$  from central fixation. The presentation of the arrays was followed by a delay period of 4,000 ms and finally the test array reappeared with a white dot appearing in the middle of the one of the items. Participants responded as to whether or not the orientation of the item with a white dot had changed. Participants completed 160 total trials with 40 trials per condition. Trials were randomly presented. Half of the trials were change trials.

### Pupil data analysis

This was the same as for Experiment 1. Missing data points due to blinks, off-screen fixations, and/or eye-tracker malfunction were removed (roughly 16.7 % of the overall data with roughly equal amounts of data loss across conditions and set sizes).

## Results and discussion

### Accuracy

First we examined accuracy as a function of each target set size (2 vs. 4) and presence of distractors (0 vs. 2). Shown in Fig. 5a is proportion correct for each condition. There was a main effect of target set size,  $F(1, 36) = 124.25$ ,  $MSE = .003$ ,  $p < .001$ , partial  $\eta^2 = .78$ , suggesting that performance decreased with larger set sizes ( $M = .90$ ,  $SE = .01$  vs  $M = .80$ ,  $SE = .02$ ). There was also a main effect of distractor presence,  $F(1, 36) = 9.00$ ,  $MSE = .002$ ,  $p = .005$ , partial  $\eta^2 = .20$ , suggesting that performance decreased when distractors were present ( $M = .86$ ,  $SE = .01$  vs  $M = .83$ ,  $SE = .02$ ). The interaction between these two was not significant,  $F(1, 36) = 2.79$ ,  $MSE = .009$ ,  $p = .104$ , partial  $\eta^2 = .07$ . Thus, consistent with prior research performance was reduced when distractors were presented and participants were required to filter them out (e.g., Robison, Miller, and Unsworth, 2017; Unsworth & Robison, 2016a; Vogel et al., 2005).

### Pupillary responses

Next we examined phasic pupillary responses as a function of target and distractor items. Phasic pupillary responses were submitted to a 2 target set size (2 vs. 4)  $\times$  2 distractor presence (0 vs. 2)  $\times$  20 (200 ms Bin) repeated measures ANOVA. There was a main effect of target set size,  $F(1, 36) = 18.61$ ,  $MSE = .045$ ,  $p < .001$ , partial  $\eta^2 = .34$ , with larger phasic responses for set size 4 compared to set size 2 ( $M = .09$ ,  $SE = .01$  vs  $M = .05$ ,  $SE = .01$ ). There was also a main effect of bin,  $F(19, 684) = 9.52$ ,  $MSE = .006$ ,  $p < .001$ , partial  $\eta^2 = .21$ . There was also a target set size  $\times$  bin interaction,  $F(19, 684) = 20.37$ ,  $MSE = .001$ ,  $p < .001$ , partial  $\eta^2 = .36$ . Consistent with the prior experiments larger set sizes demonstrated a stronger phasic response early in the delay period. Furthermore, and consistent with the prior experiments, the phasic responses associated with set size 2 increased throughout the delay period coming close to the phasic responses for set size 4 by the end of the delay period (see Fig. 5b). Finally, there was a distractor presence  $\times$  bin interaction,  $F(19, 684) = 2.70$ ,  $MSE = .001$ ,  $p < .001$ , partial  $\eta^2 = .07$ . As shown in Fig. 5b, there was no difference in the phasic responses between four red targets alone and four red targets with two blue distractors,  $F(19, 684) = 1.02$ ,  $MSE = .001$ ,  $p = .431$ , partial  $\eta^2 = .03$ . However, there were differences in the phasic responses for two red targets alone and two red targets with two blue distractors,  $F(19, 684) = 2.41$ ,  $MSE = .001$ ,  $p = .001$ , partial  $\eta^2 = .06$ . None of the other effects in the main ANOVA were significant.

These results suggest that for the most part phasic pupillary responses are tracking the number of items being maintained in WM, rather than the number of items presented. Consistent with the prior experiments phasic responses were bigger for

larger target set sizes. Additionally, when presented with four targets alone or four targets and two distractors, the phasic responses were the same suggesting that participants effectively filtered out the distractors. However, when presented with two targets alone or two targets and two distractors, the phasic response was larger for the two targets and two distractors condition. This suggests that participants were not necessarily filtering out the distractor items. Note, however, that the phasic response for this condition was smaller than the phasic response for the four targets alone condition suggesting that participants were not necessarily maintaining all four items. These differences could be due to some participants (high capacity participants) attempting to maintain all items given that the number of items presented (targets and distractors) are within their capacity. Thus, some participants might be filtering out items on the smaller target set sizes, whereas other participants decide to hold onto all items within their capacity and not do the additional work to filter out the distractors. Additionally, they could result from a subset of low capacity individuals who cannot effectively filter out the items. Overall these results are broadly consistent with prior research using the contralateral delay activity to track the number of items being maintained in WM and for tracking the ability to filter out distractors (Vogel & Machizawa, 2004, 2005).

## Experiment 5a

Experiment 4 suggested that phasic pupillary responses in a visual WM task track the ability to select target items and filter out distractors at encoding. In Experiment 5 we further examined whether these phasic responses would track selection at both encoding and during maintenance. Participants performed a version of the visual arrays task with pre-cues, retro-cues, and neutral cues (Griffin & Nobre, 2003). Prior research with this task has shown performance increases in the cued conditions compared to the neutral cue conditions suggesting that participants can select the cued item and potentially drop the other items from WM (Kuo et al., 2012; Souza et al., 2014). If this is the case then in the pre-cue condition, the results should replicate Experiment 4 showing that phasic response is smaller for the pre-cued condition than for the neutral cue condition because only one item is being maintained rather than four items. In the retro-cue condition we should see that the phasic response early in the delay period is similar to the neutral cue condition (holding four items in both cases), but following the cue the phasic response should drop to levels similar to holding only one item (i.e., the pre-cue condition). In the neutral cue condition the phasic response should continuously stay at the level of four items. This would suggest that participants can dynamically select items during both encoding and maintenance and drop other items that no

longer need to be maintained. Of course it should also be noted that there are several explanations for the retro-cue effect (see Souza & Oberauer, 2016 for a review), and thus it is possible that the retro-cue will not lead to a reduction in the phasic response compared to neutral trials.

## Method

### Participants

Participants were 44 undergraduate students recruited from the subject pool at the University of Oregon. Participants were between the ages of 18 and 35 years and received course credit for their participation. Data from ten participants were excluded from analyses because of data collection problems with the eye-tracker and data from two participants were excluded due to having accuracy of approximately 50 % or less leaving a final sample of 32 participants.

### Procedure

Participants performed a change detection task modeled after one used by Berryhill et al. (2012; see also Robison & Unsworth, 2017). Participants completed a four-item change detection task with three types of trials: neutral, pre-cue, and retro-cue. Each trial started with a 2,000-ms fixation screen with a white cross centered on a black background. After a 200-ms blank screen, the pre-cue appeared for 100 ms. On neutral and retro-cue trials, the pre-cue was a white X. On pre-cue trials, the cue was a white directional arrow pointing to one of the four locations. After a 400-ms blank screen, the memory array appeared for 300 ms. The array consisted of four colored circles. After a 2,000-ms delay, the retro-cue appeared and remained on-screen for 100 ms. On neutral and pre-cue trials, the retro-cue was a white X. On retro-cue trials, the cue was a white directional arrow pointing to one of the four locations. Cues were 100 % valid, and pre- and retro-cues did not appear on the same trials. After another 2,000-ms delay, the tested item reappeared and remained on-screen until the participant made a same/different judgment about the color of the item. Trials were presented in a random order (40 pre-cue, 40 retro-cue, and 40 neutral) for a total of 120 trials. The stimulus was equally likely to appear in each of the four locations. The color of the tested item changed on 50 % of trials.

### Pupil data analysis

This was the same as for Experiment 1. Missing data points due to blinks, off-screen fixations, and/or eye-tracker malfunction were removed (roughly 27.3 % of the overall data with roughly equal amounts of data loss across conditions and set sizes).

## Results and discussion

### Accuracy

First we examined accuracy as a function of cue type. There was an overall main effect of cue type,  $F(2, 62) = 22.12$ ,  $MSE = .002$ ,  $p < .001$ , partial  $\eta^2 = .42$ . Pre-cue trials ( $M = .95$ ,  $SE = .01$ ) trials were more accurate than neutral ( $M = .88$ ,  $SE = .02$ ),  $t(31) = 5.22$ ,  $p < .001$ ,  $d = 1.01$ , and retro-cue trials ( $M = .89$ ,  $SE = .01$ ),  $t(31) = 6.03$ ,  $p < .001$ ,  $d = 1.07$ . Interestingly, neutral and retro-cue trials were not significantly different,  $t(31) = 1.42$ ,  $p = .166$ ,  $d = .29$ , suggesting no retro-cue benefit in the current data.

### Pupillary responses

Next we examined phasic pupillary responses as a function of cue type. Phasic pupillary responses were submitted to a 3 (Cue Type: pre-cue, neutral cue, and retro-cue)  $\times$  21 (200 ms Bin) repeated measures ANOVA. There was a main effect of cue type,  $F(2, 62) = 11.21$ ,  $MSE = .082$ ,  $p < .001$ , partial  $\eta^2 = .27$ , with larger phasic responses for neutral and retro-cues than for pre-cues. There was also a main effect of bin,  $F(20, 620) = 202.54$ ,  $MSE = .029$ ,  $p < .001$ , partial  $\eta^2 = .87$ . There was also a cue type  $\times$  bin interaction,  $F(40, 1240) = 5.21$ ,  $MSE = .001$ ,  $p < .001$ , partial  $\eta^2 = .14$ . As shown in Fig. 6, neutral and retro-cues had larger phasic responses from the beginning than pre-cues, both  $F$ 's  $> 10.00$ , both  $p$ 's  $< .001$ . However, there were no differences in the phasic responses for neutral and retro-cues,  $F(1, 31) = .74$ ,  $MSE = .034$ ,  $p = .397$ , partial  $\eta^2 = .02$ . The only difference between neutral and retro-cues occurred briefly after the appearance of the cue in which retro-cues showed more constriction than the neutral cues,  $t(31) = 5.07$ ,  $p < .001$ ,  $d = .91$ . This is likely due to differences in luminance for the two cues (i.e., neutral cues were an X and retro-cues were an arrow). Furthermore, note that the reason

that the pupil response starts off negative is because there are several large changes in luminance (presentation of pre-cues, presentation of stimuli) that occur between the fixation screen where baseline is computed and the delay screen. Computing baseline as the first 200 ms of the delay period leads to overall more positive values, but does not change any of the results.

Consistent with Experiment 4 the current results suggest that when required to select items at encoding, the phasic response tracks the number of items being maintained, rather than the number of items presented. That is, pre-cue trials showed smaller phasic responses than neutral trials even though the same numbers of items were presented. These results suggest that participants were able to effectively select only the pre-cued item and maintain that item during the delay. For retro-cues the results are a little less clear. Specifically, we were unable to find a behavioral retro-cue benefit. This is surprising given that prior research has used a similar task and found robust retro-cue effects (Berryhill et al., 2012; Robison & Unsworth, 2017). However, there are a few notable differences between the tasks that could have influenced the results. Specifically, Berryhill et al. (2012) had participants engage in an articulatory suppression task to prevent verbal coding of the items. In the current version of the task we did not. Thus, it is possible that no difference between neutral and retro-cue trials could have been due to similar use of verbal rehearsal to maintain the items. Although it should be noted that we (Robison & Unsworth, 2017) used the same task as Berryhill et al. (2012) with no articulatory suppression and found robust retro-cue effects. An additional difference is that in both Berryhill et al. (2012) and Robison and Unsworth (2017) the delay period was much shorter than what was used here. The use of a much longer delay period could have changed how participants rely on the retro-cues. Given

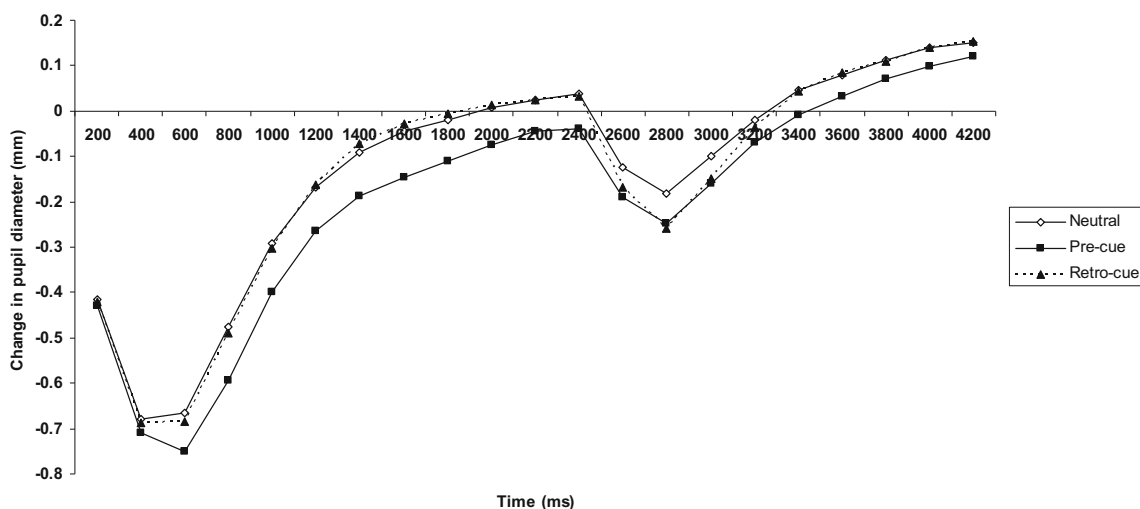
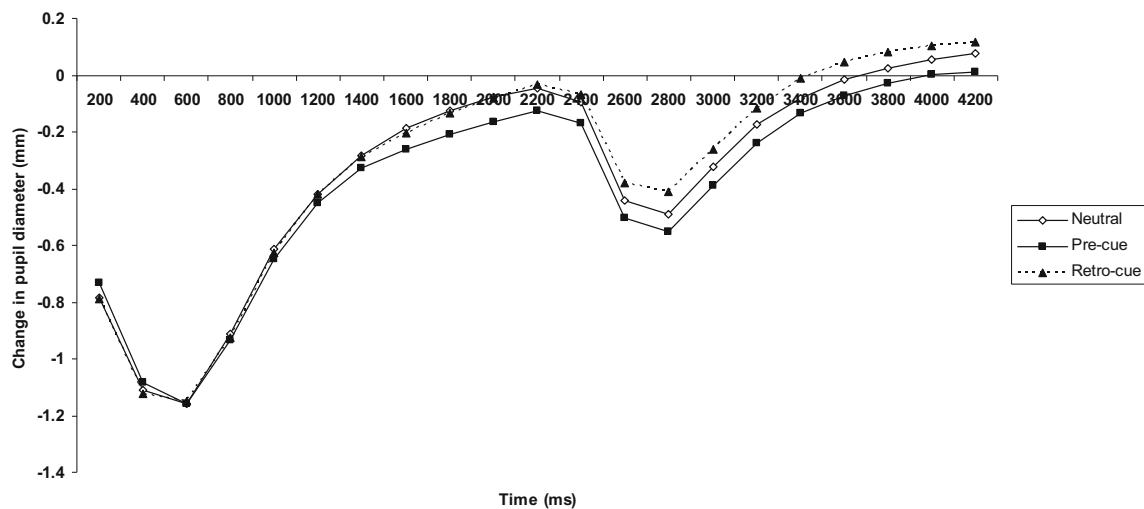


Fig. 6 Change in pupil diameter as a function of cue type and time point during the delay for Experiment 5a



**Fig. 7** Change in pupil diameter as a function of cue type and time point during the delay for Experiment 5b

that no behavioral retro-cue benefit was found, it is perhaps not surprising that the pupillary responses also showed no difference between neutral and retro-cue trials. Both the neutral and retro-cue trials demonstrated phasic responses larger than the pre-cued trials suggesting that more items were being maintained in the neutral and retro-cue trials than in the pre-cue trials.

## Experiment 5b

As noted above, an obvious limitation of Experiment 5a was the inability to find a behavioral retro-cue effect. In Experiment 5b we changed the WM task to one in which four white bars of different orientation were presented. This was done in order to attempt and replicate Experiment 5a with slightly different stimuli as well as to use stimuli that are not as easily verbalized to see if a behavioral retro-cue effect could be found.

## Method

### Participants

Participants were 40 undergraduate students recruited from the subject pool at the University of Oregon. Participants were between the ages of 18 and 35 years and received course credit for their participation. Data from seven participants were excluded from analyses because of data collection problems with the eye-tracker and data from two participants were excluded due to having accuracy of approximately 50 % or less leaving a final sample of 31 participants.

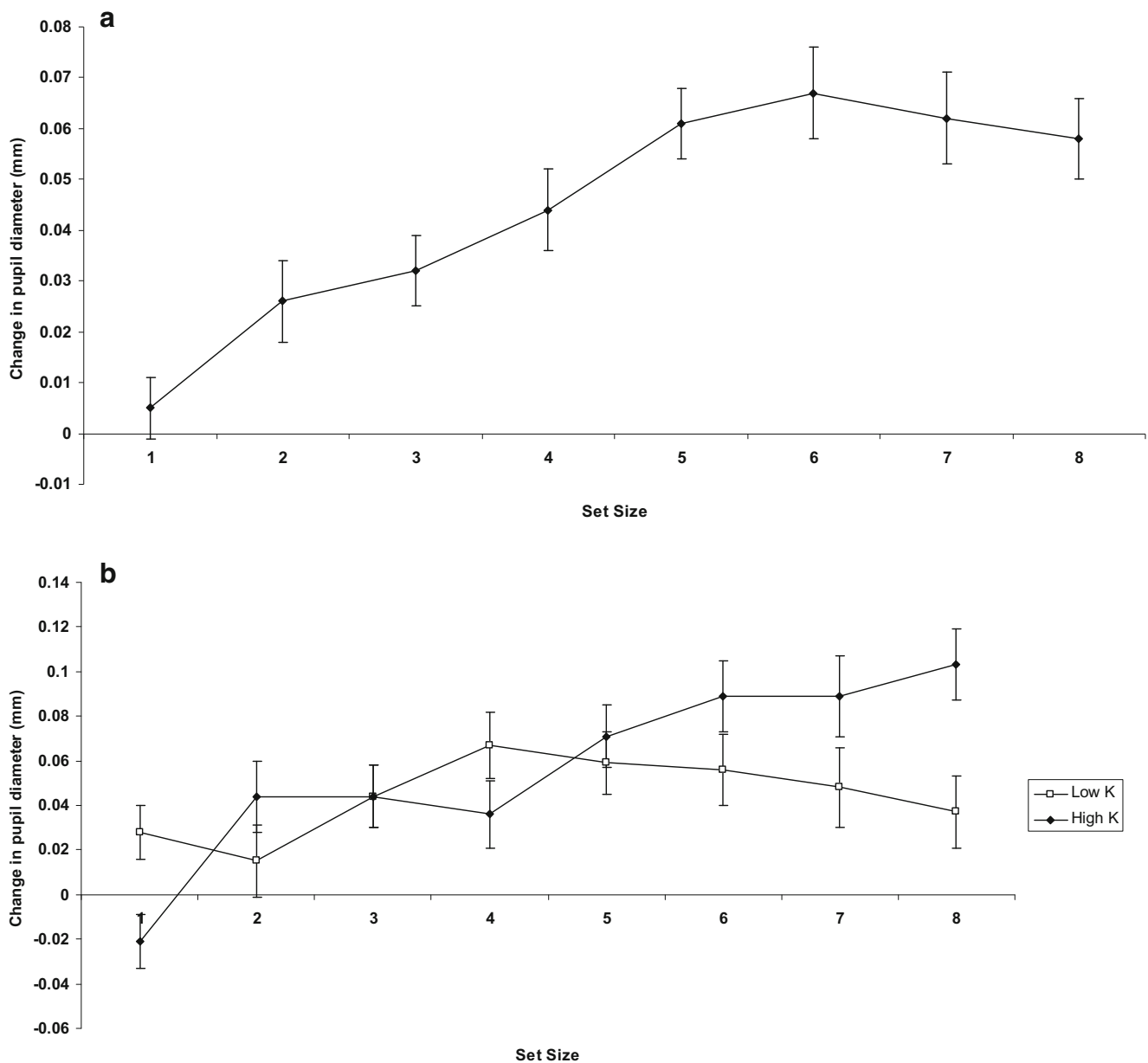
### Procedure

Participants completed a four-item change detection task with three types of trials: neutral, pre-cue, and retro-cue. Each trial started with a 2,000-ms fixation screen with a white cross centered on a black background. After a 200-ms blank screen, the pre-cue appeared for 100 ms. On neutral and retro-cue trials, the pre-cue was a white X. On pre-cue trials, the cue was a white directional arrow pointing to one of the four locations. After a 400-ms blank screen, the memory array appeared for 300 ms. The array consisted of four white bars of different orientations (vertical, horizontal, diagonal right, diagonal left). After a 2,000-ms delay, the retro-cue appeared and remained on-screen for 100 ms. On neutral and pre-cue trials, the retro-cue was a white X. On retro-cue trials, the cue was a white directional arrow pointing to one of the four locations. Cues were 100 % valid, and pre- and retro-cues did not appear on the same trials. After another 2,000-ms delay, the tested item reappeared and remained on-screen until the participant made a same/different judgment about the orientation of the item. Trials were presented in a random order (40 pre-cue, 40 retro-cue, and 40 neutral) for a total of 120 trials. The stimulus was equally likely to appear in each of the four locations. The orientation of the tested item changed on 50 % of trials.

### Pupil data analysis

This was the same as for Experiment 1. Missing data points due to blinks, off-screen fixations, and/or eye-tracker malfunction were removed (roughly 24.8 % of the overall data with roughly equal amounts of data loss across conditions and set sizes).





**Fig. 8** (a) Change in pupil diameter as function of set size in the combined analyses. (b) Change in pupil diameter as a function of set size for High and Low K individuals in the combined analyses

## Results and discussion

### Accuracy

First we examined accuracy as a function of cue type. There was an overall main effect of cue type,  $F(2, 60) = 10.99$ ,  $MSE = .004$ ,  $p < .001$ , partial  $\eta^2 = .27$ . Pre-cue trials ( $M = .89$ ,  $SE = .01$ ) were more accurate than neutral trials ( $M = .82$ ,  $SE = .02$ ),  $t(30) = 3.85$ ,  $p = .001$ ,  $d = .73$ , but were not more accurate than retro-cue trials ( $M = .88$ ,  $SE = .01$ ),  $t(30) = 1.24$ ,  $p = .23$ ,  $d = .15$ . Importantly, retro-cue trials were more accurate than neutral trials,  $t(30) = 3.46$ ,  $p = .002$ ,  $d = .75$ , suggesting the presence of a retro-cue benefit in the current data.

### Pupillary responses

Next we examined phasic pupillary responses as a function of cue type. Phasic pupillary responses were submitted to a 3 (Cue Type: pre-cue, neutral cue, and retro-cue)  $\times$  21 (200 ms Bin) repeated measures ANOVA. There was a main effect of cue type,  $F(2, 60) = 8.92$ ,  $MSE = .10$ ,  $p < .001$ , partial  $\eta^2 = .23$ , with larger phasic responses for neutral and retro-cues than for pre-cues. There was also a main effect of bin,  $F(20, 600) = 412.34$ ,  $MSE = .031$ ,  $p < .001$ , partial  $\eta^2 = .93$ . There was also a cue type  $\times$  bin interaction,  $F(40, 1200) = 9.63$ ,  $MSE = .003$ ,  $p < .001$ , partial  $\eta^2 = .24$ . As shown in Fig. 7, neutral and retro-cues had larger phasic responses than pre-cues, both  $F$ 's  $>$

6.80, both  $p$ 's  $< .02$ . However, there were no differences in the phasic responses for neutral and retro-cues,  $F(1, 30) = 2.24$ ,  $MSE = .08$ ,  $p = .145$ , partial  $\eta^2 = .07$ . The only difference between neutral and retro-cues occurred briefly after the appearance of the cue in which retro-cues showed more constriction than the neutral cues,  $t(30) = 3.40$ ,  $p = .002$ ,  $d = .61$ . Similar to Experiment 5a, this is likely due to differences in luminance for the two cues (i.e., neutral cues were an X and retro-cues were an arrow).

Overall the results from Experiment 5b were strikingly similar to Experiment 5a. Pre-cue trials were associated with higher accuracy and smaller phasic pupillary responses than neutral trials suggesting that participants were able to select and maintain one target item, rather than maintain all four target items. For retro-cue trials there was a retro-cue effect (unlike Experiment 5a) in which retro-cue trials were associated with higher accuracy than neutral trials. However, similar to Experiment 5a there were really no differences in the phasic pupillary responses for retro-cue and neutral trials. These results are inconsistent with the notion that following a retro-cue participants drop the irrelevant items from WM. If this were the case we would have expected that the following the retro-cue the phasic pupillary response would have dropped to the same levels as the pre-cue. Thus, it seems unlikely in the current data that participants are removing items from WM following a retro-cue. Rather, these results are more consistent with other accounts of retro-cue effects which suggest that the items are not dropped from WM, but rather the retro-cued items receives additional strengthening of item-context bindings, or a head start at retrieval operations (e.g., Souza & Oberauer, 2016). Future research is needed to better examine potential pupillary correlates of retro-cue effects.

### Individual differences

For our final set of analyses we examined individual differences to see if behavioral estimates of capacity (K) would correlate with the phasic pupillary responses. In prior research we found that Low K individuals had larger phasic responses for Set Size 1 than High K individuals, but the opposite was true for larger set sizes (Unsworth & Robison, 2015). Furthermore, we found that changes in the phasic response across set size (specifically the difference in the change in phasic responses from Set Size 1 to Set Size 8) correlated with behavioral estimates of capacity ( $r = .43$ ; Unsworth & Robison, 2015). To see if similar individual differences relations were present in the current results we combined data from Experiments 1–3, which used the same change detection task as Unsworth and Robison (2015). Specifically, we combined data from Experiment 1 Active condition, Experiment 2 Hold condition, Experiment 3a 4,000-ms condition, and Experiment 3b 4,000-ms condition. No participants

performed in more than one experiment. This resulted in 124 participants available for analysis. First, we examined differences in phasic responses across set sizes. There was a main effect of set size,  $F(7, 861) = 13.13$ ,  $MSE = .004$ ,  $p < .001$ , partial  $\eta^2 = .10$ . As shown in Fig. 8a, phasic pupillary responses increased from Set Size 1–5 and then plateaued. The average K estimates were 3.32 ( $SD = 1.20$ ) suggesting that the pupillary responses plateaued at a slightly higher level than the behavioral estimates of capacity. Next, we examined this as a function of individual differences with each individual's K as a covariate in an ANCOVA. Importantly, there was also a significant set size  $\times$  K interaction,  $F(7, 854) = 4.74$ ,  $MSE = .004$ ,  $p < .001$ , partial  $\eta^2 = .04$ . In order to illustrate the effects of interest we present mean changes in pupil diameter by K, via a quartile split and participants classified into Low (bottom 25 %) and High (top 25 %) K groups. Note, however, that all K analyses treated the variable as continuous, rather than as arbitrary, discrete groups. As can be seen in Fig. 8b, Low K individuals' pupil responses peaked at a lower set size than High K individuals, and High K individuals peaked at higher set sizes. Furthermore, and consistent with prior research, Low K individuals demonstrated larger pupil responses at Set size 1 compared to High K individuals (Heitz et al., 2008; Unsworth & Robison, 2015). Specifically, at Set Size 1 K was negatively correlated with the phasic pupillary response ( $r = -.23$ ), however, at Set Size 8 K was positively correlated with the phasic pupillary response ( $r = .27$ ).

We also examined changes in the phasic response as a function of set size (i.e., the difference in the change in phasic responses from Set Size 1 to Set Size 8). Consistent with prior research (Unsworth & Robison, 2015) this value was related to overall behavioral estimates of K ( $r = .47$ ).<sup>2</sup> Collectively, these results are consistent with prior research suggesting that individual

<sup>2</sup> Note, in prior research we found that not only did the change in phasic responses across set size correlate with K, but so did the coefficient of variation in pre-trial pupil diameter (Unsworth & Robison, 2015). Coefficient of variation of pre-trial pupil diameter was taken as an index of trial-to-trial fluctuations of attention. Furthermore, we found that these two were not correlated with each other and both accounted for unique variance in K. We suggested that both the amount of attention allocated to actively maintaining items in WM (phasic pupillary responses) and lapses of attention (fluctuations in pre-trial pupil diameter) accounted for individual differences in WM capacity. To see if similar results were obtained in the current data we computed the coefficient of variation for pre-trial pupil diameter in the combined data set. Consistent with prior research the coefficient of variation of pre-trial pupil diameter correlated with K ( $r = -.20$ ), but was not correlated with the phasic pupillary response ( $r = -.07$ ). Entering both values into a simultaneous regression suggested that 22% of the variance in K was accounted for and both phasic pupillary responses ( $\beta = .43$ ,  $t = 5.26$ ,  $p < .001$ ), and fluctuations in pre-trial baseline pupil diameter ( $\beta = -.18$ ,  $t = -2.16$ ,  $p = .033$ ), accounted for unique variance in K. Consistent with prior research these results suggest that individual differences in WM are due to both differences in the amount of attention that can be allocated to items (capacity) and to differences in the susceptibility to lapses of attention (Unsworth & Robison, 2015; see also Adam et al., 2015; Unsworth & Robison, 2016b).

differences in phasic pupillary responses are related to behavioral estimates of capacity.

## General discussion

In seven experiments we replicated and extended prior research examining the extent to which phasic pupillary responses track active maintenance of information in WM. The current results suggest that phasic pupillary responses track the number of items that can be maintained in WM, track when attentional effort is used to maintain items in WM, track the time course of attentional effort, and track the selection of items that are subsequently maintained in WM. Each of these are discussed in turn.

## Capacity

Consistent with prior research (Unsworth & Robison, 2015), the current results demonstrated that phasic pupillary dilations during the delay of a WM task increased and reached an asymptote around 4–5 items, suggesting a pupillary correlate of WM capacity similar to that found with contralateral delay activity (Vogel & Machizawa, 2004) and the fMRI signal in the intraparietal sulcus (Todd & Marois, 2004). This same overall pattern was seen in Experiments 1–3 using the same change detection task. Indeed, shown in Fig. 8a, are the combined results demonstrating that the phasic pupillary response increased up to around five items and then plateaued. These results suggest that phasic pupillary responses during a delay track the number of items being maintained in WM, and thus provide a physiological correlate of the capacity of WM. At the same time it is important to note that the pupillary responses sometimes plateaued at a somewhat higher level than what the behavioral estimates of capacity demonstrated. For example, in Experiment 1 the average estimate of  $K$  was 3.4 and the pupil plateaued at set size four. In the combined individual differences analyses the average estimate of  $K$  was 3.2 and the pupil plateaued at set size five. Thus, the two estimates are not exactly identical, but as suggested by the correlational analyses they are strongly related.

Examining the phasic responses across the delay suggested not only differences in the magnitude of dilation across set sizes, but also differences in the overall waveforms. Specifically, when maintaining items below capacity, the phasic response tended to be small throughout, but showing some increases towards the end of the delay (Experiments 1–4). However, when maintaining items at or above capacity, the phasic response peaked early and tended to maintain that level throughout the delay period (Experiments 1–4; at least for delays of 4,000 ms, see below on time course). As discussed below, these results are consistent with the notion that attentional effort was being allocated in a more continuous manner

to maintain the items in an active state. Importantly, the same amount of attentional effort was allocated to set sizes at or greater than capacity, suggesting that the phasic responses were tracking the number of items being maintained, rather than the number of items presented. This was true when manipulating not only set size (Experiments 1–4), but also manipulating whether or not those items had to be maintained (Experiments 1–2), and whether or not only a subset of items needed to be maintained (Experiments 4 and 5).

Furthermore, consistent with prior research, we found that individual differences in estimates of behavioral capacity were related to the pupillary estimates of capacity. Specifically, prior research found that phasic pupillary responses during the delay were related to behavioral estimates of capacity ( $r = .43$ ; Unsworth & Robison, 2015) and a similar correlation ( $r = .47$ ) was found in the current study when combining data from several experiments. As such, the current results suggest that items in WM are maintained via the continued allocation of attention effort and individual differences in the capacity of WM are partially due to differences in the amount of attention that can be allocated to actively maintain items in WM (Cowan, 2001; Craik & Levy, 1976; Unsworth & Engle, 2007). Collectively, the current results strongly suggest that phasic pupillary responses track the number of items being maintained during a delay in WM.

## Active versus passive maintenance

The current results also suggest that maintaining items in WM is an active effortful process. Specifically, in Experiment 1 participants were instructed on a trial-by-trial basis to either maintain the items (Active trials) or a later test or simply stare at the screen (Passive trials). The results suggested clear phasic pupillary responses that differed as a function of set size for the Active trials replicating prior research and the other experiments. On Passive trials, however, there was little dilation overall and no systematic effects as a function of set size, suggesting that the phasic pupillary responses were tracking the number of items that are actively being maintained in WM, rather than indexing sensory load or passive maintenance (see also Kursawe & Zimmer, 2015; and Alnaes et al., 2014 for similar results in multi-object tracking). Additionally, consistent with the notion that the phasic responses are tracking attentional effort and capacity limits of WM, the current results demonstrated similar phasic responses when maintaining only one (or two) items as when passively staring at multiple items. That is, as shown in Fig. 1b, when maintaining only one (or two) items, the phasic response demonstrated a slight early dilation followed by constriction, and then a slight increase in dilation towards the end of the delay period (see also Unsworth & Robison, 2015). As shown in Fig. 1c, similar overall phasic responses were seen when passively staring at 1–8 items, suggesting that maintaining only one (or two) items

can be done with little effort, but maintaining more items requires additional effort. It should also be noted that this differs as a function of individual differences, with low capacity individuals having to allocate more attentional effort to maintaining one item than high capacity individuals (see Fig. 8b). Thus, holding items in WM at or near one's capacity requires a great deal of continuous attentional effort.

Furthermore, in Experiment 2 participants were required to maintain items for the first half of the delay and then were given an auditory cue to either hold onto the items (Hold trials) or to drop the items (Drop trials). For the Hold trials the phasic response maintained the same level of dilation suggesting the items were still being maintained. However, on Drop trials the phasic response decreased after the drop signal and was not significantly different from baseline levels at the end of the delay. These results are consistent with prior research by Johnson (1971), in which participants performed an auditory serial recall task. In the experimental group on some trials participants heard a change in a background tone and were told that when they hear the tone change they can forget those items (directed forgetting). The control group also heard a tone, but were told that the change was meaningless. Johnson (1971) found that in the directed forgetting condition following the tone change the pupil tended to drop back to baseline levels, but in the control group the pupil continued to dilate. Thus, consistent with the prior results when told to drop items from WM, the phasic pupillary responses tend to decline back to baseline levels suggesting that they are no longer being actively maintained. Collectively, these results suggest that maintaining items in WM is an active effortful process (that occurs throughout most of the delay).

### Time course

Examining the time course of maintenance revealed several interesting findings. For example, in each experiment using a 4,000-ms delay it seemed that the phasic pupillary responses were sustained throughout the entire delay period. However, in Experiments 3a and 3b when an 8,000-ms delay was used, the phasic pupillary response declined and reached baseline levels by the end of the delay period. This suggests that participants were unable to sustain attentional effort over the entire delay period. Recently, Fabius et al. (2017) reported a similar result when examining the pupillary light response. Specifically, Fabius et al. found that the pupillary response tended to decrease after approximately 4,000 ms. Thus, the current results are consistent with similar results from a slightly different paradigm. However, as noted previously, these results are inconsistent with prior research examining multi-object tracking (which is closely related to WM; Drew & Vogel, 2008) which demonstrated that phasic pupillary responses seemed to increase up to capacity limits (and then plateau) and these phasic responses sustained throughout the

entire 10-s tracking period (Alnaes et al., 2014; Wright, Boot, & Morgan, 2013). Thus, whereas the multi-object tracking work suggests that attentional effort can be continuously allocated to objects in the environment, the current results suggest that this is not necessarily the case for items being maintained in WM in the absence of additional environmental support. Rather, the current results suggest that after 4 s or so phasic pupillary responses begin to decline indicating that attentional effort is not being allocated to the items to the same extent. As noted previously, this could occur for a number of potential reasons. It could be that items are no longer actively being maintained in WM because they have successfully been transferred to long-term memory, it could be that the phasic response is tracking decay in WM (although the accuracy results suggest otherwise), it could be that participants are switching strategies for how they are maintaining the items, or could be due to changes in how the items are being held indicating a switch to more passive maintenance. Other alternatives are also possible. None of the current results can speak to which of these possibilities are correct, but the results do suggest that participants are not sustaining active effortful processes to maintain the items in WM. Future research is needed to better understand what is occurring during these longer delays that results in a decrease in phasic responses, but does not necessarily lead to decreases in accuracy.

### Selection

In terms of selection, the current results suggested that at encoding participants could select which items to allow into and subsequently maintain in WM and that the phasic pupillary responses tracked this. Specifically, in Experiment 4 we examined whether the phasic pupillary responses would track filtering abilities in which participants were given both targets and distractors on some trials. Consistent with prior research utilizing the contralateral delay activity (Vogel et al., 2005) we found that the phasic pupillary responses tracked the number of target items being maintained rather than the number of targets and distractors that were presented (although there was evidence sometimes participants were maintaining both targets and distractors when two targets and two distractors were presented). Thus, for the most part, participants were able to filter out distractor items and only maintain target items and the phasic pupillary responses tracked this selection of items. Furthermore, in Experiments 5a and 5b, when pre-cues were used the phasic pupillary responses were smaller than when neutral cues were used suggesting that only one item was being maintained (the cued item) on pre-cue trials, whereas four items were being maintained on neutral trials. Thus, consistent with filtering, these results suggest that participants were selecting only the target item resulting in reduced pupillary responses. For retro-cues, however, the accuracy results suggested that although participants could select

the correct target item leading to a retro-cue benefit (in Experiment 5b, but not 5a), this did not lead to a change in the phasic pupillary response. Rather, the retro-cue and neutral phasic responses were nearly identical, suggesting that participants were not dropping or removing items from WM. These results are inconsistent with prior research using the CDA which demonstrated that the CDA tended to drop with retro-cues (Kuo et al., 2012). As noted previously, these results are consistent with accounts of the retro-cue benefit that suggest that the items are not removed from WM, but rather the retro-cued items receive additional strengthening of item-context bindings, or the retro-cued item gets a head start at retrieval operations (e.g., Souza & Oberauer, 2016). Future research is needed to better examine the physiological correlates of the retro-cue benefit and the possible mechanisms underlying the retro-cue benefit. Collectively, the current results suggest that phasic pupillary responses track the selection of a subset of presented items resulting in only those few items being maintained in WM rather than all items that were presented.

### Potential neural mechanisms

The current results are broadly consistent with the notion that phasic pupillary responses are tracking the number of items that are being actively maintained in WM. These results are consistent with Kahneman's (1973) suggestion that task-evoked phasic pupillary responses are linked to the intensive aspect of attention and provide an online indication of the utilization of capacity (see also Just & Carpenter, 1993). That is, the current results are consistent with the notion that maintaining items in WM is an active effortful process and that as the number of items that need to be maintained increases the amount of attention that is allocated also increases up to capacity limits.

One potential neural mechanism for interpreting the current results is the locus coeruleus norepinephrine system (LC-NE). A great deal of recent research suggests an important link between pupillary responses and the LC-NE (Alnaes et al., 2014; Aston-Jones & Cohen, 2005; Eldar, Cohen, & Niv, 2013; Gilzenrat et al., 2010; Jepma & Nieuwenhuis, 2011; Joshi et al., 2016; McGinley et al., 2015; Murphy et al., 2011; Murphy et al., 2014; Reimer et al., 2016; Samuels & Szabadi, 2008; van den Brink, Murphy, & Nieuwenhuis, 2016; Unsworth & Robison, 2016a; Varazzani et al., 2015). The LC is a brainstem neuromodulatory nucleus that is responsible for most of the NE released in the brain, and it has widespread projections throughout the neocortex including frontal-parietal areas (Berridge & Waterhouse, 2003; Samuels & Szabadi, 2008; Szabadi, 2013). Recent research suggests that the LC exhibit two general modes of firing: tonic and phasic (Aston-Jones & Cohen, 2005; Usher et al., 1999). Tonic activity refers to the overall baseline activity and phasic activity refers to the brief increase in firing rate associated with salient stimuli. In terms of pupillary responses, it is suggested that baseline pupil diameter

corresponds to LC tonic firing rate (and an overall indicator of task engagement), and task-evoked dilations correspond to LC phasic activity (and an indicator of attention allocation to task stimuli).<sup>3</sup> Thus, the phasic pupillary responses potentially reflect sustained LC-phasic mode activity in which attention is continuously allocated in order to actively maintain items in WM. Indeed, Alnaes et al. (2014) found that activity in LC (as well as the frontal eye fields and superior colliculus) correlated with pupillary changes as the number of items to be tracked increased. Thus, consistent with the current results, this suggests a potential role of the LC-NE system (specifically phasic activity) in attentional effort required to actively maintain items in WM.

Another potential neural mechanism for pupillary results seen in the current study is the superior colliculus (SC). A great deal of research suggests that the SC is important for overt and covert shifts of attention (see Krauzlis, Lovejoy, & Zenon, 2013; Corneil & Munoz, 2014 for reviews), and is a potentially important reason for the connection between WM and attention (Theeuwes, Belopolsky, & Olivers, 2009). Furthermore, recent research has shown that weak microstimulation of the SC resulted in phasic pupillary responses similar to what is seen during cognitive processing (Joshi et al., 2016; Wang et al., 2012; see also Lehmann & Corneil, 2016 for similar pupil dilation results after stimulating frontal eye fields). Wang and Munoz (2015) further suggested that "SC-mediated pupil pathways could provide the substrate required for pupil size modulation by various cognitive processes" (p. 139). Indeed, recent research has found that the pupil (in particular the pupillary light reflex) can be used to track covert shifts of attention during WM maintenance (Fabius et al., 2017; Unsworth & Robison, 2017). Additionally, as noted above, Alnaes et al. (2014) found that SC activity correlated with pupillary changes as the number of items to be tracked increased. Thus, it is possible that the pupillary responses seen in the current study are due to covert shifts of attention to selected items in WM mediated via SC activity. This notion is consistent with the attention-based rehearsal hypothesis which suggests that items are maintained in WM via covert shifts of attention to prioritized information during WM maintenance (Awh & Jonides, 2001; Awh, Vogel, & Oh, 2006). Thus, the current pupillary results could reflect not only LC-NE mediated attention allocation as more items need to be maintained in WM, but also SC mediated covert shifts of attention to prioritized locations in order to use attention-based rehearsal processes.

<sup>3</sup> Although prior work is suggestive of a link between LC-NE functioning and pupil diameter, it should be acknowledged that this relation is correlational in nature and it could be due to the fact that both the LC and the sympathetic nervous system are linked via a third system such as the nucleus paragigantocellularis (Nieuwenhuis, de Geus, & Aston-Jones, 2011).

## Conclusions

The current study examined pupillary correlates of WM maintenance. In particular, the results suggested the phasic pupillary responses increased as the number of items that needed to be maintained increased up to around 4–5 items consistent with behavioral estimates of capacity. These phasic pupillary responses were related to capacity estimates at an individual differences level. Furthermore, the phasic pupillary responses demonstrated WM load dependent relations only when the items needed to be actively maintained. When told to passively stare at the screen or drop the current items, the pupil remained near baseline levels. The phasic pupillary responses also tracked the time course of maintenance demonstrating sustained responses for the first 4,000 ms, but declines thereafter. Finally, the phasic pupillary responses provided evidence for selection processes at encoding in terms of tracking both filtering abilities and the effect of pre-cues. Collectively these results suggest that phasic pupillary responses can be used to track the active maintenance of items in WM.

**Author Notes** This research was supported by Office of Naval Research grant N00014-15-1-2790.

## References

- Adam, K.C.S., Mance, I., Fukuda, K., & Vogel, E.K. (2015). The contributions of attentional lapses to individual differences in working memory capacity. *Journal of Cognitive Neuroscience*, *8*, 1601–1616.
- Allen, R.J., Baddeley, A.D., Hitch, G.J. (2017). Executive and perceptual distraction in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *43*, 1677–1693.
- Alnaes, D., Sneve, M. H., Espeseth, T., Endestad, T., van de Pavert, S. H. P., & Laeng, B. (2014). Pupil size signals mental effort deployed during multiple object tracking and predicts brain activity in the dorsal attention network and the locus coeruleus. *Journal of Vision*, *14*, 1–20.
- Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus–norepinephrine function: Adaptive gain and optimal performance. *Annual Review of Neuroscience*, *28*, 403–450.
- Awh, E., & Jonides, J. (2001). Overlapping mechanisms of attention and working memory. *Trends in Cognitive Sciences*, *5*, 119–126.
- Awh, E., Vogel, E., & Oh, S.-H. (2006). Interactions between attention and working memory. *Neuroscience*, *139*, 201–208.
- Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin*, *91*, 276–292.
- Beatty, J., Lucero-Wagoner, B., 2000. The pupillary system. In: Cacioppo, J.T., Tassinary, L.G., Berntson, G.G. (Eds.), *Handbook of Psychophysiology* (pp. 142–162). Cambridge University Press, New York.
- Belopolsky, A.V. & Theeuwes, J. (2009). No functional role of attention-based rehearsal in maintenance of spatial working memory representations. *Acta Psychologica*, *132*, 124–135.
- Berridge C.W., Waterhouse B.D. (2003) The locus coeruleus–noradrenergic system: Modulation of behavioral state and state-dependent cognitive processes. *Brain Research Reviews*, *42*, 33–84.
- Berryhill, M. E., Richmond, L. L., Shay, C. S., & Olson, I. R. (2012). Shifting attention among working memory representations: Testing cue type, awareness, and strategic control. *The Quarterly Journal of Experimental Psychology*, *65*, 426–438.
- Corneil, B. D. and Munoz, D. P. (2014) Overt responses during covert orienting. *Neuron*, *82*, 1230–1243.
- Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*, 97–185.
- Cowan, N. (2005). Working memory capacity. Hove: Psychology Press.
- Cowan, N, Elliott, E.M., Saults, J.S., Morey, C.C., Mattox, S., Hismjatullina, A., & Conway, A.R.A. (2005). On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. *Cognitive Psychology*, *51*, 42–100.
- Craik, F.I.M., & Levy, B.A. (1976). The concept of primary memory. In W.K. Estes (Ed.), *Handbook of Learning and Cognitive Processes* (pp. 133–175). New York: Lawrence Erlbaum Associates.
- Daniels, L. B., Nichols, D. F., Seifert, M. S., and Hock, H. S. (2012). Changes in pupil diameter entrained by cortically initiated changes in attention. *Vision Neuroscience*, *29*, 131–142.
- Drew, T. & Vogel, E. K. (2008). Neural measures of individual differences in selecting and tracking multiple moving objects. *Journal of Neuroscience*, *28*, 4183–4191.
- Eldar, E., Cohen, J.D., & Niv, Y. (2013). The effects of neural gain on attention and learning – *Nature Neuroscience*, *16*, 1146–1153.
- Engle, R. W., & Kane, M. J. (2004). Executive attention, working memory capacity, and a two-factor theory of cognitive control. In B. Ross (Ed.), *The psychology of learning and motivation* (Vol. 44, pp. 145–199). NY: Elsevier.
- Engle, R.W., Tuholski, S.W., Laughlin, J.E., & Conway, A.R.A. (1999). Working memory, short-term memory and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General*, *128*, 309–331.
- Fabius, J.H., Mathôt, S., Schut, M.J., Nijboer, T.C.W., & Van der Stigchel, S. (2017). Focus of spatial attention during spatial working memory maintenance: Evidence from pupillary light response. *Visual Cognition*, *25*, 10–20.
- Fougnie, D. (2009). The relationship between attention and working memory. In N. B. Johansen (Ed.), *New research on short-term memory*. NY: Nova Science Publishers.
- Fougnie, D., & Marois, R. (2006). Distinct capacity limits for attention and working memory. Evidence from attentive tracking and visual working memory paradigms. *Psychological Science*, *17*, 526–534.
- Fukuda, K. Vogel, E.K., Mayr, U., & Awh, E. (2010). Quantity not quality: The relationship between fluid intelligence and working memory capacity. *Psychonomic Bulletin & Review*, *17*, 673–679.
- Gilzenrat, M. S., Nieuwenhuis, S., Jepma, M., & Cohen, J. D. (2010). Pupil diameter tracks changes in control state predicted by the adaptive gain theory of locus coeruleus function. *Cognitive, Affective, & Behavioral Neuroscience*, *10*, 252–269.
- Gold, J. M., Wilk, C. M., McMahon, R. P., Buchanan, R. W., & Luck, S. J. (2003). Working memory for visual features and conjunctions in schizophrenia. *Journal of Abnormal Psychology*, *112*, 61–71.
- Goldinger, S.D., & Papesch, M.H. (2012). Pupil dilation reflects the creation and retrieval of memories. *Current Directions in Psychological Science*, *21*, 90–95.
- Griffin, I. C., & Nobre, A. C. (2003). Orienting attention to locations in internal representations. *Journal of Cognitive Neuroscience*, *15*, 1176–1194.
- Heitz, R. P., Schrock, J. C., Payne, T. W., & Engle, R. W. (2008). Effects of incentive on working memory capacity: Behavioral and pupillometric data. *Psychophysiology*, *45*, 119–129.
- Hess, E. H., & Polt, J. M. (1964). Pupil size in relation to mental activity during simple problem-solving. *Science*, *143*, 1190–1192.
- Hollingworth, A., & Maxcey-Richard, A. M. (2013). Selective maintenance in visual working memory does not require sustained visual

- attention. *Journal of Experimental Psychology: Human Perception and Performance*, 39, 1047–1058.
- Jepma, M., & Nieuwenhuis, S. (2011). Pupil diameter predicts changes in the exploration-exploitation trade-off: Evidence for the adaptive gain theory. *Journal of Cognitive Neuroscience*, 23, 1587–1596.
- Johnson, D.A. (1971). Pupillary responses during a short-term memory task: Cognitive processing, arousal, or both? *Journal of Experimental Psychology*, 90, 311–318.
- Johnson, E.L., Miller Singley, A.T., Peckham, A., Johnson, S., & Bunge, S.A. (2014). Task evoked pupillometry provides a window into the development of short-term memory capacity. *Frontiers in Developmental Psychology*, 5, 218.
- Joshi, S., Li, Y., Kalwani, R.M., & Gold, J.I. (2016). Relationship between pupil diameter and neuronal activity in the locus coeruleus, colliculi, and cingulate cortex. *Neuron*, 89, 221–234.
- Just, M.A., & Carpenter, P.A. (1993). The intensity dimension of thought: Pupillometric indices of sentence processing. *Canadian Journal of Experimental Psychology*, 47, 310–339.
- Kahneman, D. (1973). *Attention and effort*. Upper Saddle River: Prentice Hall.
- Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. *Science*, 154, 1583–1585.
- Krauzlis, R.J., Lovejoy, L.P., & Zenon, A. (2013). Superior colliculus and visual spatial attention. *Annual Review of Neuroscience*, 36, 165–182.
- Kuo, B.C., Stokes, M.G., Nobre, A.C. (2012) Attention modulates maintenance of representations in visual short-term memory. *Journal of Cognitive Neuroscience*, 24, 51–60.
- Kursawe, M. A., Zimmer, H. D. (2015). Costs of storing color and complex shape in visual working memory: Insights from pupil size and slow waves. *Acta Psychologica*, 158, 67–77.
- Laeng, B., Sirois, S., & Gredebäck, G. (2012). Pupillometry: A Window to the Preconscious?. *Perspectives on Psychological Science*, 7, 18–27.
- Lee, E., Cowan, N., Vogel, E.K., Rolan, T., Valle-Inclán, F., & Hackley, S.A. (2010). Visual working memory deficits in Parkinson's patients are due to both reduced storage capacity and impaired ability to filter out irrelevant information. *Brain*, 133, 2677–2689.
- Lehmann, S.J., & Corneil, B.D. (2016). Transient pupil dilation after subsaccadic microstimulation of primate frontal eye fields. *Journal of Neuroscience*, 36, 3765–3776.
- Luck, S. J. & Vogel, E. K. (2013). Visual working memory capacity: From psychophysics and neurobiology to individual differences. *Trends in Cognitive Science*, 17, 391–400.
- McGinley, M.J., David SV, & McCormick, D.A. (2015). Cortical membrane potential signature of optimal states for sensory signal detection. *Neuron*, 87, 179–192.
- Morey, C. C., & Bieler, M. (2013). Visual short-term memory always requires general attention. *Psychonomic Bulletin & Review*, 20, 163–170.
- Morey, C.C., & Cowan, N. (2004). When visual and verbal memories compete: Evidence of cross-domain limits in working memory. *Psychonomic Bulletin & Review*, 11, 296–301.
- Morey, C.C., & Cowan, N. (2005). When do visual and verbal memories conflict? The importance of working-memory load and retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 703–713.
- Murphy P.R., O'Connell R.G., O'Sullivan M., Robertson I.H., Balsters J.H. (2014) Pupil diameter covaries with BOLD activity in human locus coeruleus. *Human Brain Mapping* 35, 4140–4154.
- Murphy, P.R., Robertson, I.H., Balsters, J.H. & O'Connell, R.G. (2011). Pupillometry and P3 index the locus coeruleus-noradrenergic arousal function in humans. *Psychophysiology*, 48, 1532–43.
- Naber, M., Alvarez, G.A., & Nakayama, K. (2013). Tracking the allocation of attention using human pupillary oscillations. *Frontiers in Psychology*, 4, 1–12.
- Nieuwenhuis, S., de Geus, E.J., & Aston-Jones, G. (2011). The anatomical and functional relationship between the P3 and autonomic components of the orienting response. *Psychophysiology*, 48, 162–175.
- Parra MA, Abrahams S, Logie RH, Mendez LG, Lopera F, Della SS. (2010). Visual short-term memory binding deficits in familial Alzheimer's disease. *Brain*, 133, 2702–2713.
- Peavler, W.S. (1974). Pupil size, information overload, and performance differences. *Psychophysiology*, 11, 559–566.
- Reimer, J., McGinley, M.J., Liu, Y., Rodenkirch, C., Wang, Q., McCormick, D.A., & Tolia, A.S. (2016). Pupil fluctuations track rapid changes in adrenergic and cholinergic activity in cortex. *Nature Communications*, 7, 13289.
- Robison, M. K., Miller, A. L., & Unsworth, N. (2017). Individual differences in working memory capacity and filtering. *Journal of Experimental Psychology: Human Perception and Performance*, (in press).
- Robison, M.K., & Unsworth, N. (2017). Variation in the use of cues to guide attention in visual working memory. *Attention, Perception, & Psychophysics*, 79, 1652–1665.
- Samuels, E.R., Szabadi, E. (2008). Functional neuroanatomy of the noradrenergic locus coeruleus: Its roles in the regulation of arousal and autonomic function. Part I: Principles of functional organization. *Current Neuropharmacology*, 6, 235–253.
- Shipstead, Z., Lindsey, D.R.B., Marshall, R.L., Engle, R.W. (2014). The mechanisms of working memory capacity: Primary memory, secondary memory, and attention control. *Journal of Memory and Language*, 72, 116–141.
- Souza, A.S., & Oberauer, K. (2016). In search of the focus of attention in working memory; 13 years of the retro-cue effect. *Attention, Perception, & Psychophysics*, 78, 1839–1860.
- Souza, A. S., Jerker, L., & Oberauer, K. (2014). Unloading and reloading working memory: Attending to one item frees capacity. *Journal of Experimental Psychology: Human Perception and Performance*, 40, 1237–1256.
- Szabadi, E. (2013). Functional neuroanatomy of the central noradrenergic system. *Journal of Psychopharmacology*, 27, 659–693.
- Theeuwes, J., Belopolsky, A. & Olivers, C.N.L. (2009). Interactions between working memory, attention and eye movements. *Acta Psychologica*, 132, 106–114.
- Todd, J. J. & Marois, R. (2004). Capacity limit of visual short-term memory in human posterior parietal cortex. *Nature*, 428, 751–754.
- Todd, J. J., & Marois, R. (2005). Posterior parietal cortex activity predicts individual differences in visual short-term memory capacity. *Cognitive, Affective, and Behavioral Neuroscience*, 5, 144–155.
- Tsukahara, J.S., Harrison, T.L., & Engle, R.W. (2016). The relationship between baseline pupil size and intelligence. *Cognitive Psychology*, 91, 109–123.
- Unsworth, N., & Engle, R.W. (2007). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review*, 114, 104–132.
- Unsworth, N, Fukuda, K., Awh, E., & Vogel, E.K. (2015). Working memory delay activity predicts individual differences in cognitive abilities. *Journal of Cognitive Neuroscience*, 27, 853–865.
- Unsworth, N., Fukuda, K., Awh, E., & Vogel, E. K. (2014). Working memory and fluid intelligence: Capacity, attention control, and secondary memory retrieval. *Cognitive Psychology*, 71, 1–26.
- Unsworth, N. & Robison, M.K. (2015). Individual differences in the allocation of attention to items in working memory: Evidence from pupillometry. *Psychonomic Bulletin & Review*, 3, 757–765.
- Unsworth, N. & Robison, M.K. (2016a). Pupillary correlates of lapses of sustained attention. *Cognitive, Affective, & Behavioral Neuroscience*, 16, 601–615.
- Unsworth, N., & Robison, M.K. (2016b). The influence of lapses of attention on working memory capacity. *Memory & Cognition*, 44, 188–196.

- Unsworth, N., & Robison, M.K. (2017). Pupillary correlates of covert shifts of attention during working memory maintenance. *Attention, Perception, & Psychophysics*, *79*, 782-795.
- Usher, M., Cohen, J. D., Servan-Schreiber, D., Rajkowski, J., & Aston-Jones, G. (1999). The role of locus coeruleus in the regulation of cognitive performance. *Science*, *283*, 549-554.
- van den Brink, R.L., Murphy, P.R., & Nieuwenhuis, S. (2016). Pupil diameter tracks lapses of attention. *PLoS ONE 11*: e0165274.
- Varazzani, C., San-Galli, A., Dilardeau, S., & Bouret, S. (2015). Noradrenaline and dopamine neurons in the reward/effort trade-off: A direct electrophysiological comparison in behaving monkeys. *Journal of Neuroscience*, *35*, 7866-7877.
- Vogel, E. K. & Machizawa, M. G. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature*, *428*, 784-751.
- Vogel, E. K., McCollough, A. W., & Machizawa, M. G. (2005). Neural measures reveal individual differences in controlling access to visual working memory. *Nature*, *438*, 500-503.
- Wang, C.A., Boehnke, S.E., White, B.J., and Munoz, D. P. (2012) Microstimulation of the monkey superior colliculus induces pupil dilation without evoking saccades. *Journal of Neuroscience*, *32*, 3629-3636.
- Wang C-A, Munoz D.P. (2015) A circuit for pupil orienting responses: Implications for cognitive modulation of pupil size. *Current Opinion in Neurobiology*, *33*, 134-140.
- Wright, T. J., Boot, W. R., & Morgan, C. S. (2013). Pupillary response predicts multiple object tracking load, error rate, and conscientiousness, but not inattentive blindness. *Acta Psychologica*, *144*, 6-11.