Perturbing Neural Representations of Working Memory with Task-irrelevant Interruption

Nicole Hakim[1], Tobias Feldmann-Wüstefeld[2], Edward Awh[1], and Edward K. Vogel[1]

Abstract

■ Working memory maintains information so that it can be used in complex cognitive tasks. A key challenge for this system is to maintain relevant information in the face of task-irrelevant perturbations. Across two experiments, we investigated the impact of task-irrelevant interruptions on neural representations of working memory. We recorded EEG activity in humans while they performed a working memory task. On a subset of trials, we interrupted participants with salient but task-irrelevant objects. To track the impact of these task-irrelevant interruptions on neural representations of working memory, we measured two well-characterized, temporally sensitive EEG markers that reflect active, prioritized working memory representations: the contralateral delay activity and lateralized alpha power (8–12 Hz). After interruption, we found that contralateral delay activity amplitude momentarily sustained but was gone by the end of the trial. Lateralized alpha power was immediately influenced by the interrupters but recovered by the end of the trial. This suggests that dissociable neural processes contribute to the maintenance of working memory information and that brief irrelevant onsets disrupt two distinct online aspects of working memory. In addition, we found that task expectancy modulated the timing and magnitude of how these two neural signals responded to task-irrelevant interruptions, suggesting that the brain's response to task-irrelevant interruption is shaped by task context. ■

INTRODUCTION

Working memory is a large-scale neural system that maintains readily accessible task-relevant information via active neural firing. A key challenge for this system is to protect these active representations from task-irrelevant interruptions. Extensive prior work has characterized how the presence of irrelevant information during the encoding of targets (distractors) impacts working memory representations (Feldmann-Wüstefeld & Vogel, 2018; Gaspar & McDonald, 2014; Clapp, Rubens, & Gazzaley, 2010; Postle, D'Esposito, & Corkin, 2005). This work has revealed that the presence of distractors during this initial encoding period (0–500 msec) greatly reduces working memory performance, in part because these items compete with targets for limited representational space in working memory (Olivers, 2008; Vogel, McCollough, & Machizawa, 2005; De Fockert, Rees, Frith, & Lavie, 2001). After this initial encoding period, presence of irrelevant information (interrupters) has a reduced but still measurable impact on performance (Vogel, Woodman, & Luck, 2006). These interrupters have less of an impact because working memory representations have reached a more stable state, which is consistent with the time course of neural measures of working memory representations (Ikkai, McCollough, & Vogel, 2010; Vogel et al., 2006). This reduced impact is also likely because of the formation of concurrent visual long-term memory representations that represent the targets in a passive yet still accessible format (e.g., Fukuda & Vogel, 2019; Chun & Turk-Browne, 2007; Woodman & Chun, 2006). Together, these concurrent active and passive representations of targets reduce the behavioral impact of interruption during working memory maintenance. Yet, despite robust behavioral performance, current models of working memory still predict that onsets of task-irrelevant interruption should produce a momentary perturbation of the maintained target representations during which attention is withdrawn from the targets and at least temporarily applied to the positions of the interrupters (Bisley, Zaksas, Droll, & Pasternak, 2004; Bisley & Goldberg, 2003). However, the consequences of such a brief withdrawal of attention on the neural signatures of working memory are not well understood. Here, we seek to measure the impact that task-irrelevant interruption has on the ongoing active neural representations of targets held in working memory.

To track the impact of task-irrelevant interruption on neural representations of working memory, we measured two well-characterized, temporally sensitive EEG markers that reflect active, prioritized working memory representations: the contralateral delay activity (CDA) and lateralized alpha power (8–12 Hz). The CDA is a sustained negative-going wave in human EEG that tracks current working memory load. It is sensitive to trial-by-trial fluctuations in working memory performance and distinguishes stable individual differences in working memory (Luria, Balaban,

[1]University of Chicago, [2]University of Southampton

Journal of Cognitive Neuroscience 32:3, pp. 558–569
https://doi.org/10.1162/jocn_a_01481

Awh, & Vogel, 2016; Vogel & Machizawa, 2004). This component is thought to reflect an index of the current items that are actively represented in working memory (Hakim, Adam, Gunseli, Awh, & Vogel, 2019; Feldmann-Wüstefeld, Vogel, & Awh, 2018). Lateralized alpha power is similarly sensitive to task-relevant information. This signal is measured as a decrease in alpha power over posterior electrodes that are contralateral to the position of the attended items. However, despite its similarity to the CDA, it has been shown to be a distinct component of actively maintained information (Fukuda, Mance, & Vogel, 2015) that appears to primarily track the current position of spatial attention (Foster, Bsales, Jaffe, & Awh, 2017; Foster, Sutterer, Serences, Vogel, & Awh, 2016; Thut, Nietzel, Brandt, & Pascual-Leone, 2006; Worden, Foxe, Wang, & Simpson, 2000). Topographic distributions of alpha power across the entire scalp have been shown to contain precise spatial information about remembered/attended stimuli (van Moorselaar et al., 2018; Foster et al., 2017), whereas lateralized alpha power has been used as an effective tool for establishing which visual hemifield is currently attended. Together, the CDA and lateralized alpha power respectively provide an item-based and space-based index of task-relevant information that is actively represented in working memory. Furthermore, because both signals are lateralized, we were able to isolate processing of the memory array by presenting the memory items laterally and the interrupters along the vertical midline of the display. As items on the vertical midline do not affect lateralized signals, the ongoing lateral measures only reflect processing of the memory representations. This allowed us to measure how these working memory representations respond to task-irrelevant interruption.

In Experiment 1, we sought to determine how task-irrelevant interrupters impact ongoing working memory representations. We did this by presenting midline interrupters during the retention interval of a working memory task. In Experiment 2, we sought to determine whether the neural responses to task-irrelevant interrupters could be modulated by task expectancy. During all of these tasks, we recorded EEG activity from human participants.

EXPERIMENT 1

Methods

Participants

Twenty-two volunteers naive to the objective of the experiment participated for payment (~15 USD per hour). All data were collected in a single session. The data of two participants were excluded from the analysis because of too many artifacts, poor behavioral performance (see below for criteria), or technical problems. The remaining 20 participants (12 men) were between the ages of 19 and 30 years (M = 22.7, SD = 3.4). Participants in all experiments reported normal or corrected-to-normal visual acuity as well as normal color vision. All experiments were conducted with the written understanding and consent of each participant. The University of Chicago Institutional Review Board approved experimental procedures.

Stimuli

All stimuli were presented on a gray background (~33.3 cd/m^2). Cue displays showed a small central fixation dot (0.2° × 0.2°). A horizontal diamond composed of a green triangle (RGB = 74, 183, 72; 52.8 cd/m^2) and a pink triangle (RGB = 183, 73, 177; 31.7 cd/m^2) appeared on the vertical midline 0.65° above the fixation dot. In 50% of the trials, the pink triangle pointed to the left side and the green triangle pointed to the right side; in the remaining 50% of the trials, this was inverse. Half of the participants were instructed to attend the hemifield that the pink triangle pointed to, and the other half was instructed to attend the hemifield the green triangle pointed to. Memory displays showed a series of colored squares (1.1° × 1.1°, mean luminance = 43.1 cd/m^2). Colors for the squares were selected randomly from a set of 11 possible colors (red = 255, 0, 0; green = 0, 255, 0; blue = 0, 0, 255; yellow = 255, 255, 0; magenta = 255, 0, 255; cyan = 0, 255, 255; purple = 102, 0, 102; brown = 102, 51, 0; orange = 255, 128, 0; white = 255, 255, 255; black = 0, 0, 0). Squares could appear within an area of the display subtending 3.5° to the left or right of fixation and 3.1° above and below fixation. There was the same number of squares in each hemisphere. Within each hemisphere, squares were as equally distributed between the upper and lower hemifields as possible. The interruption display showed four colored squares of the same size as the memory display, drawn from the remaining colors. These interrupting items were shown on the vertical midline with a randomly jittered horizontal offset of maximally 0.55° (half of an object). Retention interval displays were blank with a small central fixation dot (0.2° × 0.2°). Probe displays showed one colored square in each hemisphere in the same location as one of the squares in the original array. In 50% of the trials, the color was identical (no change trials) to the memory display. In the remaining 50% of the trials, it was one of the colors not used in the memory or interruption display (change trials). The same stimuli were used in all experiments.

Apparatus

Participants were seated with a chin rest in a comfortable chair in a dimly lit, electrically shielded and sound-attenuated chamber. Participants responded with button presses on a standard keyboard that was placed in front of them. Stimuli were presented on an LCD computer screen (BenQ XL2430T; 120-Hz refresh rate, 61-cm screen size in diameter; 1920 × 1080 pixels) placed at a 74-cm distance from participants. An IBM-compatible

computer (Dell Optiplex 9020) controlled stimulus presentation and response collection.

Procedure

Each trial began with a cue display (500 msec) indicating the relevant side of the screen (left or right). A memory display consisting of six colored squares in each hemifield followed the cue display for 150 msec. Participants were instructed to memorize as many colored squares in the memory display from the cued side and to ignore the other side entirely, as that side would never be probed. Participants had to remember the items for a retention interval of 1650 msec during which a central fixation dot was shown. In 25% of the trials, an interruption display appeared 500 msec after memory display offset for 150 msec. The total length of the retention interval was 1650 msec, regardless of whether an interruption appeared. Participants were instructed to always ignore interruption displays. After the retention interval, a probe display appeared until response. Participants had to indicate whether the color at the probed location changed color ("?/" key) or did not change color ("z" key). After participants responded, the trial concluded, and the next trial started after a blank intertrial interval of 750 msec. Participants completed 1200 trials (15 blocks of 80 trials), that is, 300 trials with interruption and 900 trials without interruption. Information about average performance and a minimum break of 30 sec were provided after each block. See Figure 1 for a visual depiction of the task.

We presented the interrupters in locations that did not overlap with the locations of the memory items to avoid visual masking. Importantly, the relative position of interrupters and targets matters in lateralized change detection tasks. When interrupters are presented laterally with targets on the vertical midline, the neural signature of sustained interrupters suppression can be isolated (Contralateral Delay Activity, positive). Conversely, when interrupters are presented on the vertical midline and targets are presented laterally, the neural signature of target processing can be isolated (Feldmann-Wüstefeld & Vogel, 2018). Accordingly, because we were interested in how neural representations of targets are affected by interruption, we placed the interrupters along the vertical midline. Thus, reductions in CDA amplitude can be interpreted as dropping memory items, and reductions in lateralized alpha power can be interpreted as a shift of attention away from the laterally presented memory arrays.

Behavioral Data Analysis

We separately analyzed performance for the trials with and without interruption. Performance was converted to a capacity score, K, calculated as $N \times (H - FA)$, where N is the set size, H is the hit rate, and FA is the false alarm rate (Cowan, 2011). To compare performance between the two conditions, we used a two-tailed, repeated-measures t test.

Artifact Rejection

We recorded EEG activity from 30 active Ag/AgCl electrodes (Brain Products actiCHamp) mounted in an elastic cap positioned according to the International 10–20 system (Fp1, Fp2, F7, F8, F3, F4, Fz, FC5, FC6, FC1, FC2, C3, C4, Cz, CP5, CP6, CP1, CP2, P7, P8, P3, P4, Pz, PO7, PO8, PO3, PO4, O1, O2, Oz). FPz served as the ground electrode, and all electrodes were referenced on-line to TP10 and rereferenced off-line to the average of all electrodes. Incoming data were filtered (low cutoff = 0.01 Hz, high cutoff = 80 Hz, slope from low to high cutoff = 12 dB/octave) and recorded with a 500-Hz sampling rate. Impedances were kept below 10 kΩ. To identify trials that were contaminated with eye movements and blinks, we used EOG activity and eye tracking. We collected EOG data with five passive Ag/AgCl electrodes (two vertical EOG electrodes placed above and below

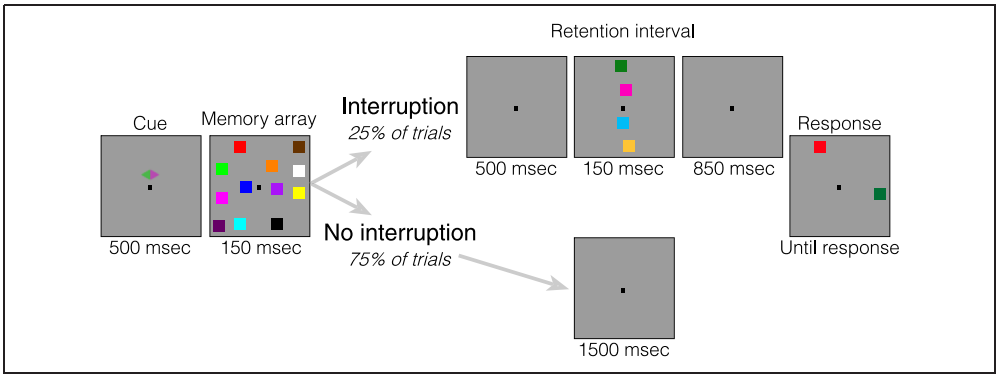

Figure 1. Task for Experiment 1. At the start of each trial, a cue appeared on the screen for 500 msec, which cued participants to attend one side of the screen. Then, an array of four colored squares briefly appeared (150 msec). On 75% of trials (no interruption condition), the retention interval (1500 msec) remained blank the entire time. On the other 25% of trials (interruption condition), the retention interval was blank for 500 msec, but then a series of four colored squares appeared on the midline for 150 msec. Participants were told to always ignore these squares that appeared on the midline of the screen during the retention interval. After the brief interruption, the screen then went blank again for 850 msec. At the end of each trial, a response screen appeared with one square in each hemifield. Participants were told to report whether the square on the attended side was the same color as the original square in that location.

the right eye, two horizontal EOG (HEOG) electrodes placed ~1 cm from the outer canthi, and one ground electrode placed on the left cheek). We collected eye-tracking data using a desk-mounted EyeLink 1000 Plus eye-tracking camera (SR Research Ltd.) sampling at 1000 Hz. Usable eye-tracking data were acquired for 20 of 22 participants in Experiment 1 and 29 of 30 participants in Experiment 2.

EEG was segmented off-line with segments time-locked to memory display onset. Eye movements, blinks, blocking, drift, and muscle artifacts were first detected by applying automatic detection criteria to each segment. After automatic detection (see below), trials were manually inspected to confirm that detection thresholds were working as expected. Participants were excluded if they had less than 100 correct trials remaining in any of the conditions. For the participants used in analyses, we rejected, on average, 21% of trials in Experiment 1 and 39% of trials in Experiment 2.

Eye Movements

We used a sliding window step function to check for eye movements in the HEOG and the eye-tracking gaze coordinates. For HEOG rejection, we used a split-half sliding window approach. We slid a 100-msec time window in steps of 10 msec from the beginning to the end of the trial. If the change in voltage from the first half to the second half of the window was greater than 20 μV, it was marked as an eye movement and rejected. For eye-tracking rejection, we applied a sliding window analysis to the *x*-gaze coordinates and *y*-gaze coordinates (window size = 100 msec, step size = 10 msec, threshold = 0.5° of visual angle).

Blinks

We used a sliding window step function to check for blinks in the vertical EOG (window size = 80 msec, step size = 10 msec, threshold = 30 μV). We checked the eye-tracking data for trial segments with missing data points (no position data are recorded when the eye is closed).

Drift, Muscle Artifacts, and Blocking

We checked for drift (e.g., skin potentials) by comparing the absolute change in voltage from the first quarter of the trial to the last quarter of the trial. If the change in voltage exceeded 100 μV, the trial was rejected for drift. In addition to slow drift, we checked for sudden step-like changes in voltage with a sliding window (window size = 100 msec, step size = 10 msec, threshold = 100 μV). We excluded trials for muscle artifacts if any electrode had peak-to-peak amplitude greater than 200 μV within a 15-msec time window. We excluded trials for blocking if any electrode had at least 30 time points in any given 200-msec time window that were within 1 V of each other.

CDA Analysis

Segmented EEG data were baselined from 200 to 0 msec before the onset of the memory displays. Artifact-free EEG segments from correct trials were averaged separately for each condition (no interruption, interruption) and separately for electrodes ipsilateral and contralateral to the attended side. Then, the difference between contralateral and ipsilateral activity for the electrode pair PO7/PO8 was calculated (i.e., the CDA), resulting in two average waveforms for each participant. The average CDA amplitude was calculated for three time windows: before interruption onset (450–650 msec), after interruption onset (800–1000 msec), and before probe onset (1300–1500 msec). We then compared the CDA for each time window with a repeated-measures two-tailed *t* test. To measure the robustness of the CDA for each condition (reliable difference between contralateral and ipsilateral activity), we also ran a one-sample *t* test (against zero) for each time window.

Lateralized Alpha Power Analysis

The same EEG segments as the CDA analysis were used in this analysis; however, the segments were not baselined. The raw EEG signal was band-pass filtered in the alpha band (8–12 Hz) using a two-way least-squares finite-impulse-response filter (*eegfilt.m* from EEGLAB Toolbox). Instantaneous power was then extracted by applying a Hilbert transform (*hilbert.m*) to the filtered data. The resulting data were averaged separately for each condition (no interruption, interruption) and each laterality (contralateral vs. ipsilateral to cued hemifield) for the electrode pair PO7/PO8, resulting in four average waveforms for each participant. The average alpha power was calculated for the same three time windows as the CDA analysis. We then compared lateralized alpha power suppression for each time window with a repeated-measures two-tailed *t* test. To measure the robustness of lateralized alpha power suppression for each condition (reliable difference between contralateral and ipsilateral activity), we also ran a one-sample *t* test (against zero) for each time window.

Results

Behavior

Performance (Figure 2), as measured by *K*, was significantly worse on trials that were interrupted ($M = 1.6$) than on trials that were not interrupted ($M = 1.2$), using a significant two-way repeated-measures *t* test, $t(19) = 4.428$, $p < .001$, $M = 0.408$, 95% CI [0.215, 0.601].

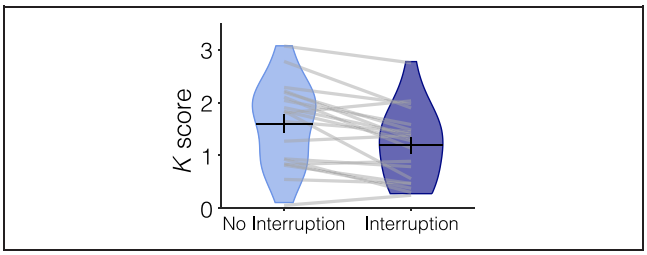

Figure 2. Behavioral performance for Experiment 1 separated by condition. Participants remembered fewer items when they were interrupted (dark blue plot) than when they were not interrupted (light blue plot). Average performance (*K* score) is represented with the horizontal black line and the black error bars reflect the *SEM*. The distribution of *K* scores for all participants is represented by the violin plot. Light gray lines connect data from one participant across conditions.

CDA

Pre-interruption (450–650 msec). The CDA (Figure 3) was robust before interruption onset (450–650 msec) on trials with, $t(19) = -3.187$, $p = .005$, $M = -0.707$, 95% CI [−1.171, −0.243], and without, $t(19) = -4.053$, $p = .001$, $M = -0.837$, 95% CI [−1.270, −0.405], interruption. There was not a significant difference in CDA amplitude between trials with and without interruption during this time window, $t(19) = -1.394$, $p = .179$, $M = -0.131$, 95% CI [−0.327, 0.066].

Post-interruption 1 (800–1000 msec). Immediately after the offset of the interruption (800–1000 msec), the CDA remained robust for both conditions, namely, no interruption: $t(19) = -4.016$, $p = .001$, $M = -0.674$, 95% CI [−1.025, −0.323]; interruption: $t(19) = -2.928$, $p = .009$, $M = -0.583$, 95% CI [−1.000, −0.166]. Again, there was not a significant difference in CDA amplitude between trials with and without interruption, $t(19) = -0.525$, $p = .606$, $M = -0.091$, 95% CI [−0.452 0.271].

Post-interruption 2 (1300–1500 msec). By the end of the trial (1300−1500 msec), however, there was a significant difference in CDA amplitude between trials with and without interruption, $t(19) = -5.145$, $p \leq .001$, $M = -0.731$, 95% CI [−1.028, −0.434]. On trials without interruption, the CDA remained robust, $t(19) = -2.535$, $p = .020$, $M = -0.496$, 95% CI [−0.906, −0.086]. However, on trials with interruption, the CDA was no longer significantly different from zero, 1.445, $p = .165$, $M = 0.234$, 95% CI [−0.105, 0.574].

Lateralized Alpha Power

Pre-interruption (450–650 msec). Alpha power (Figure 3) was significantly more negative at contralateral compared with ipsilateral electrodes before interruption onset (450–650 msec) on trials with, $t(19) = -2.131$, $p = .046$, $M = -7.264$, 95% CI [−14.398, −0.130], and without, $t(19) = -2.517$, $p = .021$, $M = -8.815$, 95% CI [−16.145, −1.486], interruption. Alpha power was significantly more lateralized on trials without interruption than trials with interruption during this time window, $t(19) = -2.573$, $p = .019$, $M = -1.551$, 95% CI [−2.813, −0.289]. We suspect that this may be because of time smearing in the alpha analysis. Time smearing is a side effect of Fourier transforms, as the calculation of power at any time point incorporates data from time points before and after the time point of interest. Therefore, the effect of the interruption may be smeared

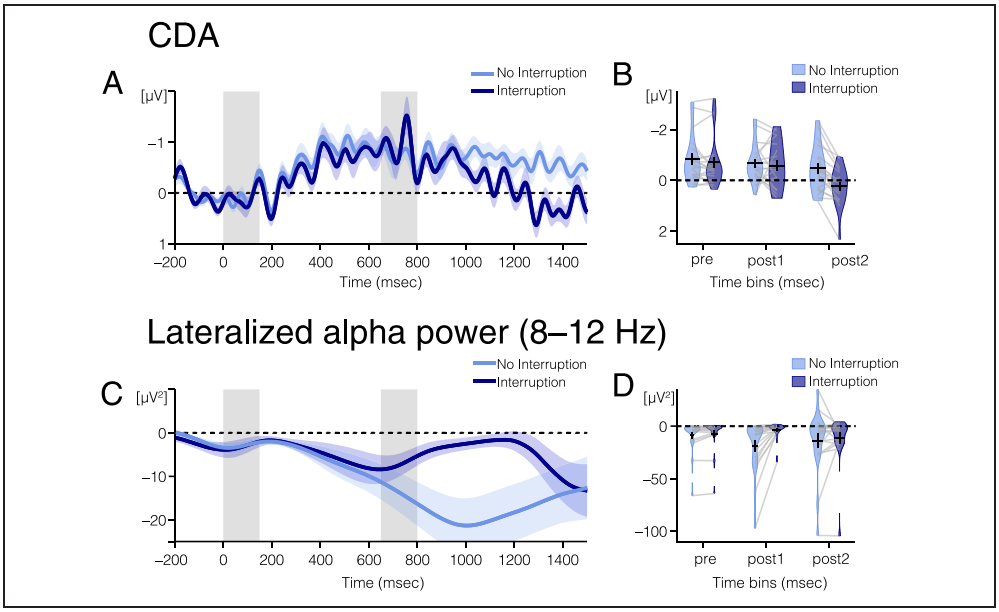

Figure 3. EEG results from Experiment 1. (A) CDA amplitude and (C) alpha power lateralization over time for trials with (dark blue line) and without (light blue line) interrupters. The light color envelopes around each line represent *SEM* for each condition. The first vertical gray bar (time point: 0–150 msec) represents when the memory array was on the screen, and the second gray bar (time point: 650–800 msec) represents when the interrupters were on the screen, if there were interrupters on that trial. (B) CDA amplitude and (D) alpha power lateralization for trials with (light blue plots) and without (dark blue plots) interruption averaged over the three time windows of interest (pre: 450–650; post1: 800–1000; and post2: 1300–1500 msec). (B) Average CDA amplitude and (D) average alpha power lateralization are represented with the horizontal black line. The black error bars reflect the *SEM*. The colored area of the violin plots reflects the distribution of amplitudes for all participants. Light gray lines connect data from one participant across conditions.

across time, causing it to appear like there are differences in alpha power before interruption onset when there actually are only differences after interruption onset.

Post-interruption 1 (800–1000 msec). Immediately after the offset of the interruption (800–1000 msec), lateralization of alpha power remained robust after trials without interruption, $t(19) = -3.423$, $p = .003$, $M = -19.393$, 95% CI [−31.250, −7.536]. However, alpha power was not significantly lateralized after trials with interruption, but this effect was trending toward significance, $t(19) = -2.054$, $p = .054$, $M = -3.554$, 95% CI [−7.175, 0.067]. During this time window, lateralized alpha power was significantly more lateralized on trials without interruption than trials with interruption, $t(19) = -3.629$, $p = .002$, $M = -15.839$, 95% CI [−24.974, −6.704].

Post-interruption 2 (1300–1500 msec). By the end of the trial (1300−1500 msec), however, there was no longer a significant difference in lateralized alpha power between the two conditions, $t(19) = -0.904$, $p = .377$, $M = -3.879$, 95% CI [−12.863, 5.104]. Lateralized alpha power was significantly lateralized in both conditions, namely, no interruption: $t(19) = -2.144$, $p = .045$, $M = -14.207$, 95% CI [−28.077, −0.337]; interruption: $t(19) = -2.137$, $p = .046$, $M = -10.328$, 95% CI [−20.444, −0.213].

Conclusions

In Experiment 1, participants' working memory performance was reduced when they were interrupted during the retention interval. In addition, both the CDA and lateralized alpha power were negatively impacted by the interrupters, but this effect had distinct time courses for the two signals. The CDA briefly sustained after interruption, whereas alpha power immediately became less lateralized. By the end of the trial, CDA was no longer present, but alpha power relateralized. These results suggest that the CDA and lateralized alpha power respond distinctly to task-irrelevant interruptions.

EXPERIMENT 2

In Experiment 2, we sought to determine whether the neural responses to task-irrelevant interruptions could be modulated by task expectancy or if they are fixed responses to interruption irrespective of the participants' expectations. Thus, in Experiment 2, we compared the same 25% interruption condition employed in Experiment 1 with one in which interrupters were presented on 75% of trials. We predicted that a higher frequency of task-irrelevant interruptions should allow participants to be better prepared for interruptions. Accordingly, CDA and lateralized alpha power should sustain longer after interruption. Importantly, we will also examine the onset and offset of the CDA and lateralized alpha power to examine whether the time course of the two subprocesses may be affected differently.

Methods

Participants

Thirty novel volunteers naive to the objective of the experiment participated for payment (~15 USD per hour). All data were collected in a single session. The data of 10 participants were excluded from the analysis because of too many artifacts, poor behavioral performance, or technical problems (same criteria as in Experiment 1). The remaining 20 participants (11 men) were between the ages of 19 and 32 years ($M = 23.54$, $SD = 3.85$).

Stimuli and Procedures

Stimuli were identical to Experiment 1 (Figure 1). Procedure was also identical to Experiment 1, except for the following changes. The retention interval was increased to 2000 msec. The experiment was divided in two halves in each of which the probability of interruption was varied. The order of the halves was counterbalanced across participants. The probability for interruption was 25% in one part (no interruption: 75%) and 75% in the other part (no interruption: 25%). This resulted in 2 × 2 design with the factors Interruption (no interruption vs. interruption) and Probability (high vs. low). Participants completed 1920 trials in total (24 blocks of 80 trials each), 240 trials for each of the two low-probability conditions and 720 trials for each of the two high-probability conditions.

Analysis

Behavioral and EEG data were analyzed analogously to Experiment 1 but included the additional factor Probability. For statistical analyses, we forwarded the mean CDA amplitude (contralateral minus ipsilateral activity) to a two-way ANOVA with the within-participant factors Interruption (interruption vs. no interruption) and Probability (75% probability for interruption vs. 25%). In addition, for the CDA and lateralized alpha analyses, the time window before probe onset was 1800–2000 msec, as we extended the retention interval by 500 msec.

Results

Behavior

Performance (Figure 4), as measured by K, was significantly worse on trials that were interrupted (low probability: $M = 1.4$, high probability: $M = 1.6$) than on trials that were not interrupted (low probability: $M = 1.6$, high probability: $M = 1.7$), regardless of probability, with a significant main effect of Interruption, $F(1, 19) = 21.288$,

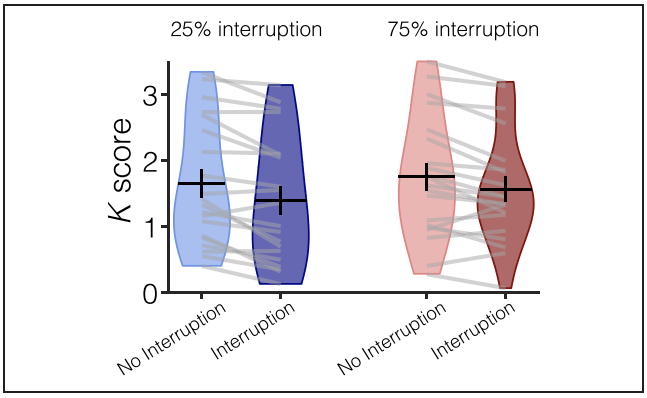

Figure 4. Behavioral performance for Experiment 2 separated by condition. Participants remembered fewer items when they were interrupted (dark blue plot) than when they were not interrupted (light blue plot). Average performance (*K* score) is represented with the horizontal black line and the black error bars reflect the *SEM*. The distribution of *K* scores for all participants is represented by the violin plot. Light gray lines connect data from one participant across conditions.

$p < .001$, $\eta_p^2 = .528$. There was not a significant main effect of Probability, $F(1, 19) = 3.575$, $p = .074$, $\eta_p^2 = .158$, or an interaction of Interruption and Probability, $F(1, 19) = 1.420$, $p = .248$, $\eta_p^2 = .070$.

CDA

Pre-interruption 2 (450–650 msec). Before interruption onset (450–650 msec), there was a significant CDA in all four conditions (all one-sample t tests: $p \leq .002$). In addition, there was no difference in CDA amplitude between any of the conditions, $p \geq .060$, for the main effects of Interruption, Probability, and their interaction, although the interaction of Interruption and Probability was trending toward significance, $p = .060$, $\eta_p^2 = .173$.

Post-interruption 1 (800–1000 msec). Immediately after interruption offset (800–1000 msec), the influence of interruption on CDA amplitude depended on the probability of being interrupted, with a significant interaction of Probability and Interruption, $F(1, 19) = 9.951$, $p = .005$, $\eta_p^2 = .344$. Follow-up t tests run separately for trials with and without interruption revealed that, when interrupters were present, the amplitude of the CDA depended on the probability of interruption, $t(19) = 2.252$, $p = .036$. The CDA was significantly larger in the high-probability condition ($M = -0.660$) than in the low-probability condition ($M = -0.265$). On trials without interruption, there was no difference in CDA amplitude between probabilities, $t(19) = -0.858$, $p = .402$. The main effects of Interruption and Probability were not significant, $p \geq .129$.

Post-interruption 2 (1800–2000 msec). By the end of the trial (1800–2000 msec), there was no difference in CDA amplitude between any of the conditions, $p \geq .142$, for the main effects of Interruption, Probability, and their interaction. There was no longer a significant CDA in any condition (all one-way t tests: $p \geq .088$). CDA tends to decline over time, and by extending the delay period compared with Experiment 1, we may have reached the point at which the CDA tends to decline naturally (Vogel & Machizawa, 2004).

Lateralized Alpha Power

Pre-interruption 2 (450–650 msec). Alpha power (Figure 5) was significantly suppressed in all conditions before the interruption onset (450–650 msec; all one-sample t tests: $p \leq .005$). The influence of interruption on alpha power suppression depended on the probability of interruption, with a significant interaction of Probability and Interruption, $F(1, 19) = 4.881$, $p = .040$, $\eta_p^2 = .204$. As in Experiment 1, this pre-interruption difference could be because of time smearing of alpha power. There was no difference in lateralized alpha power between trials that were and were not interrupted for either the high-probability, $F(1, 19) = -2.046$, $p = .055$, or low-probability, $F(1, 19) = 0.279$, $p = .789$, trials, although this effect was trending in the high-probability trials.

Post-interruption 1 (800–1000 msec). Immediately after the offset of interruption (800–1000 msec), alpha power was significantly lateralized in all conditions (all one-sample t tests: $p \leq .006$). There was a significant main effect of Interruption on the strength of alpha lateralization, $F(1, 19) = 6.530$, $p = .019$, $\eta_p^2 = .256$. For both high- and low-probability trials, lateralized alpha power was stronger on trials without interruption (low probability: $M = -15.498$, high probability: $M = -15.368$) than on trials with interruption (low probability: $M = -4.549$, high probability: $M = -4.966$). No other effects were significant, $p \geq .778$.

Post-interruption 2 (1800–2000 msec). By the end of the trial (1800–2000 msec), the influence of interruption on lateralized alpha power depended on the probability of interruption, with a significant interaction of Interruption and Probability, $F(1, 19) = 6.365$, $p = .021$, $\eta_p^2 = .251$. The difference between lateralized alpha power in high- and low-probability trials is different with and without interruption. However, within high-probability trials, follow-up t tests revealed that there was not an alpha power suppression difference between trials that were interrupted and those that were not for either high-probability, $t(19) = -1.923$, $p = .070$, or low-probability, $t(19) = 1.077$, $p = .295$, trials. In addition, alpha power suppression was not significantly different between high- and low-probability trials for trials that were interrupted, $t(19) = -1.278$, $p = .217$, or for those that were not interrupted, $t(19) = 1.362$, $p = .189$. The interaction of Interruption and Probability is disordinal because the interaction is significant, but the follow-up t tests are not significant. Disordinal interactions indicate that a factor has one kind of effect in one condition and the

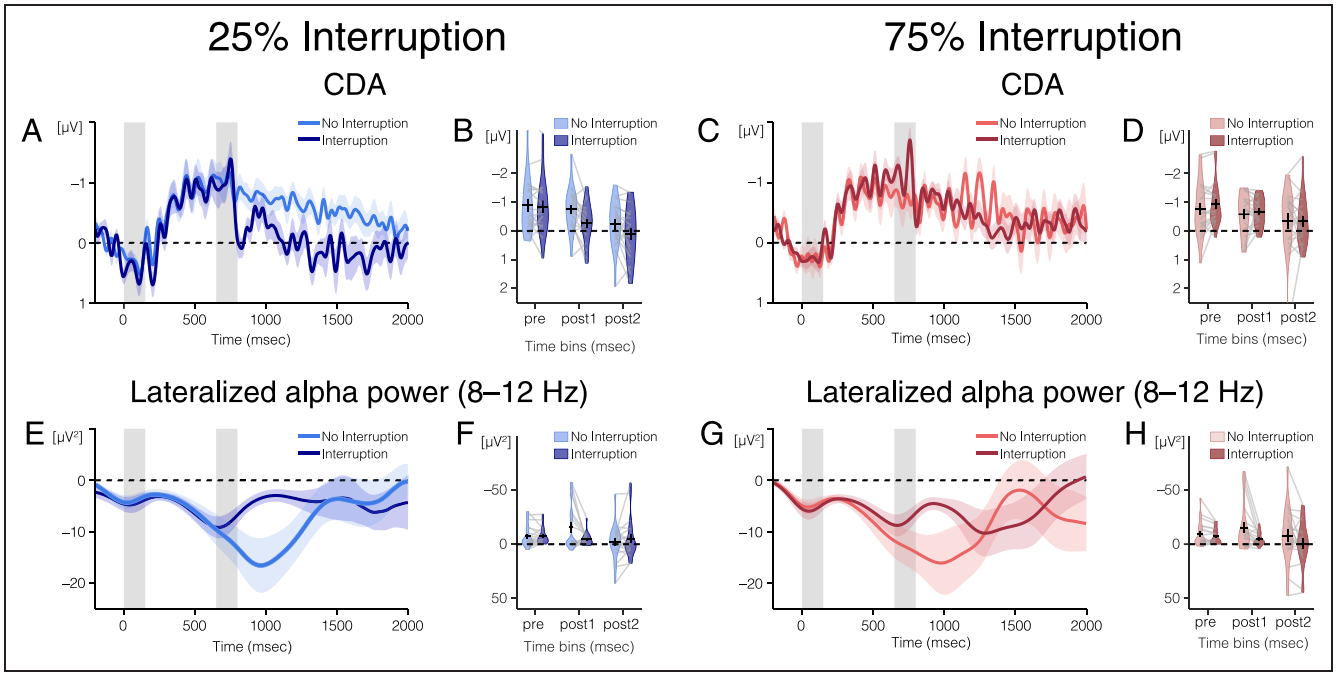

Figure 5. EEG results from Experiment 2. (A) CDA amplitude and (E) alpha power lateralization over time for trials without interruption in the 25% (light blue) and 75% (pink) interruption blocks. The light color envelopes around each line represent *SEM* for each condition. The first vertical gray bar (time points: 0–150 msec) represents when the memory array was on the screen, and the second gray bar (time points: 650–800 msec) represents when the interrupters were on the screen, if there were interrupters on that trial. (B) CDA amplitude and (F) alpha power lateralization for trials without interruption in the 25% (light blue plots) and 75% (pink plots) interruption blocks averaged over the three time windows of interest (450–650, 800–1000, and 1800–2000 msec). (B) Average CDA amplitude and (F) average alpha power lateralization are represented with the horizontal black lines and the black error bars reflect the *SEM*. The colored area of the violin plots reflects the distribution of (B) CDA amplitudes and (F) alpha power lateralization for all participants. Light gray lines connect data from one participant across conditions. (C) CDA amplitude and (G) lateralized alpha power over time for trials with interruption in the 25% (dark blue) and 75% (red) interruption blocks. (D) CDA amplitude and (H) alpha power lateralization averaged over three time windows of interest for trials with interruption in the 25% (dark blue plot) and 75% (red plot) interruption blocks.

opposite kind of effect in the other condition. In this case, in the low-probability condition, alpha power is numerically more lateralized when interrupters were present ($M = -5.14$, $SD = 18.89$) than when they were not present ($M = -1.84$, $SD = 16.53$). However, in the high-probability condition, alpha power is numerically less lateralized when interrupters were present ($M = -0.58$, $SD = 20.83$) than when they were not present ($M = -7.296$, $SD = 25.92$). The main effects of Interruption and Probability were also not significant, $p \geq .472$.

Conclusions

In Experiment 2, we found that behavioral performance was worse when participants were interrupted than when they were not interrupted regardless of the probability of interruption. Once again, the neural results revealed that both CDA and lateralized alpha power were negatively impacted by interruptions, but these two signals had distinct time courses. After interruption, CDA sustained, but lateralized alpha power became less lateralized. In addition, the amplitude of CDA immediately after interruption depended on the probability of interruption—CDA amplitude was larger when participants were expecting to be interrupted. However, alpha power lateralization did not depend on expectations—alpha power shifted toward baseline when interruptions were present, regardless of the probability of interruption.

By the end of the trial, CDA was no longer present on trials with interruptions, regardless of probability. This replicates the CDA results from Experiment 1. The effect of probability and interruption on alpha power lateralization by the end of the trial was a bit more ambiguous. In the low-probability block, lateralized alpha power was equivalent on trials with and without interruptions. This replicates the results from Experiment 1. However, upon visual inspection of the results, the "recovery" pattern after interruption was not as apparent. This is because overall alpha power lateralization on trials without interruption was very close to baseline, unlike in Experiment 1 where alpha power was robustly lateralized. This reduction in alpha power lateralization on trials without interruption could plausibly be because of the length of the retention interval. In Experiment 1, the retention interval was 1500 msec, and in Experiment 2, it was extended to 2000 msec. We did this so that we could investigate whether CDA would return if participants had more time post-interruption. However, both CDA and alpha power lateralization tend to shift

toward baseline with longer delays, which may be the reason why alpha power is less lateralized by the end of the trial in Experiment 2 than in Experiment 1. Regardless of the amount of alpha power lateralization at the end of the trial, we still found a significant reversal of the effect of probability and interruption in the high-probability condition as compared with the low-probability condition. In the high-probability condition, alpha power was more lateralized on trials without interruptions than on trials with interruptions, but this effect was reversed in the low-probability condition.

GENERAL DISCUSSION

Working memory maintains information so that it can be used despite momentary perturbations from task-irrelevant information. Here, we examined how memory representations that have already reached a stable state respond to visual interruption. As expected, we found a modest behavioral impact of interruption. Participants remembered significantly fewer items when they were interrupted than when they were not interrupted, but they performed above chance in all conditions. Despite a modest behavioral impact, task-irrelevant interruption produced substantial perturbations on two well-characterized EEG signals of working memory, lateralized alpha power and CDA. Both lateralized alpha power, an index of sustained spatial attention, and CDA, an index of actively maintained working memory representations, were disrupted at certain points during the delay, but the time course of these perturbations varied. Lateralized alpha power results suggest that attention shifted toward baseline immediately after the interruption but had returned to the target positions by the end of the trial. By contrast, the CDA results suggest that working memory representations continued to persist after the interruption but was eliminated by the end of the trial. We additionally found that task expectancy modulated the timing and magnitude of these perturbations of working memory representations, suggesting that the brain's response to task-irrelevant interruption is regulated by task context. The distinct time courses of and the influence of task context on lateralized alpha power and CDA have many interesting theoretical implications that future work can help elucidate.

Neural Response Immediately After Interruption

Sudden onsets of task-irrelevant interruption have been shown to capture attention when interrupters are visually salient (van Moorselaar et al., 2018; Andrews, Ratwani, & Trafton, 2009; Bisley & Goldberg, 2003). In our experiment, we used lateralized alpha power as an index of sustained spatial attention (Hakim et al., 2019; Foster et al., 2016). After the onset of interruption, lateralized alpha power almost immediately shifted toward baseline. When lateralized alpha power is at baseline, it suggests that participants are no longer spatially attending the lateral memory items. Neural evidence from previous studies suggests that participants attend to the locations of interrupting stimuli (van Moorselaar et al., 2018; Bisley & Goldberg, 2003) because of attentional capture (Feldmann-Wüstefeld & Schubö, 2013; Sawaki & Luck, 2012). Thus, in this study, participants presumably shifted their attention away from lateralized representations after the onset of task-irrelevant interruption to the centrally presented interrupters.

During this same time window, CDA remained robust and significantly above baseline. Previous research has shown that CDA is sensitive to trial-by-trial fluctuations in working memory performance and tracks the number of maintained object representations (Adam, Mance, Fukuda, & Vogel, 2015; Ikkai et al., 2010). Considering this, the robust CDA immediately after the onset of interruption suggests that object representations persist, at least momentarily, after the withdrawal of spatial attention to a new position. The presence of CDA and lack of lateralized alpha power immediately after interruption raise the long-standing theoretical question of whether object representations maintained in working memory can persist without sustained spatial attention. Previous research has suggested that spatial attention is a rehearsal mechanism that facilitates the maintenance of object representations held in working memory (Williams & Woodman, 2012). In addition, the positions of object representations are maintained in working memory even when spatial information is completely irrelevant (Foster et al., 2017). Together, these previous results suggest that spatial attention aids the maintenance of working memory information but do not address whether working memory representations necessitate sustained spatial attention. In this study, the robust CDA and lack of lateralized alpha power after the onset of interruption suggest that object representations maintained in working memory can persist without sustained spatial attention. Therefore, our results suggest that working memory representations may not necessitate sustained spatial attention. Nevertheless, working memory representations may still be volatile without sustained spatial attention, given that CDA goes to baseline by the end of trials with interruption.

Neural Activity at the End of Interrupted Trials

In this study, we sought to interrupt participants after working memories reached a stable state. Therefore, it is not surprising that participants can still perform well above chance in the distractor-present trials. It is likely that interruptions to the working memory representations at earlier moments, such as before CDA is fully formed, would produce larger behavioral decrements (e.g., Vogel et al., 2006). Nevertheless, by the end of interrupted trials, we observed that the CDA was no longer reliable, but alpha power became relateralized. There is a

large body of research that has shown that CDA tracks the active maintenance of information, is sensitive to trial-by-trial fluctuations in working memory performance, and distinguishes stable individual differences in working memory (Luria et al., 2016; Vogel & Machizawa, 2004). Therefore, the pattern of activity at the end of the trial suggests that participants reoriented their attention to the locations of the memoranda but no longer maintained active working memory representations. If participants no longer maintained object representations that are tracked by CDA, how were they able to still perform the change detection task on interrupted trials (albeit worse than noninterrupted trials)? There are a few possible explanations.

One possible explanation for the absence of the CDA at the end of the trial but above-chance behavioral performance is that performance on interrupted trials could rely on offline memory representations. Previous research has shown that information in working memory can be simultaneously maintained in both active and passive memory states (Mallett & Lewis-Peacock, 2018). Therefore, when actively maintained memory traces are no longer present, information could still be retrieved from an offline state. Research that has investigated retrieval of information from offline memory states has found that alpha power tracks information retrieved from long-term memory (Fukuda, Kang, & Woodman, 2016). In addition, other research has suggested that attention can aid recall of information that would be otherwise unavailable to working memory (Murray, Nobre, Clark, Cravo, & Stokes, 2013). These findings dovetail with our results—at the end of interrupted trials, when information about the memoranda is required to respond to the probe, lateralized alpha power could be reinstated to reload information from offline memory storage, thereby bolstering behavioral performance. An alternate explanation for the recovery of lateralized alpha power at the end of the trial is that it reflects the anticipation of the upcoming memory probe. The memory probe always appeared in the same location as one of the memory items. Thus, to shift attention to the location of the upcoming probe, participants had to remember the locations of the original memory items. Therefore, even if the relateralization of alpha power at the end of interrupted trials reflects the orienting of spatial attention to the location of the anticipated memory probe, it still suggests that this relateralization relies on the retrieval of task-relevant spatial information. Both the reloading and reorienting explanations of the recovery of alpha power are plausible and theoretically interesting explanations for this pattern of activity.

The relatively good behavioral performance without CDA could alternately be explained by other neural traces of actively maintained working memory representations that we are not measuring. The CDA is a coarse neural measure that compares activity contralateral and ipsilateral to memory items. Thus, it is not an exhaustive measure of working memory. More spatially global neural signals or more distributed patterns of activity, for example, could sustain after task-irrelevant interruption, and these signals could plausibly bolster behavioral performance. Regardless of the mechanism that preserves information about the memoranda, our results strongly suggest that actively maintained information is dynamically perturbed after task-irrelevant interruption.

Modulation of CDA and Alpha Power by Task Demands

In Experiment 2, we varied task demands by interrupting participants on 75% (high) or 25% (low) of trials. After interruption onset, we found the same pattern of result as Experiment 1; lateralized alpha power shifted toward baseline while CDA persisted. However, the amplitude of the CDA varied as a function of task demands. When task demands were high, CDA amplitude was higher than when task demands were low. This suggests that participants were able to better protect working memory representations when they were expecting to be interrupted and that task context is involved in how the brain responds to task-irrelevant interruption. On the other hand, the influence of task demands on lateralization of alpha power was more ambiguous. Our results suggest that spatial attention may be uniformly captured by interrupters initially regardless of expectation. However, during certain points in the trial, lateralization of alpha power may vary as a function of task demands. Therefore, the neural responses to interruption that we observed were affected both by both interruption and task demands. These results go hand in hand with previous research that has shown that distraction by salient irrelevant stimuli can be modulated by top–down control. For example, when a color singleton is presented on 20% of the trials, it slows down RTs in a visual search task more than when it is presented on 50% of trials (Folk & Remington, 2015; Marini, Chelazzi, & Maravita, 2013; Müller, Geyer, Zehetleitner, & Krummenacher, 2009; Horstmann, 2005) because attention requires more time to be deployed to the relevant information when rare distractors appear (Töllner, Müller, & Zehetleitner, 2012).

Conclusions

In this set of experiments, we investigated the impact of task-irrelevant interruption on two dissociable neural signals, namely, CDA, a neural index of actively maintained representations, and lateralized alpha power, an index of sustained spatial attention. By tracking these neural markers of working memory, we were able to observe changes in active representations that would not be apparent from behavioral measures alone. Both CDA and lateralized alpha power were impacted by task-irrelevant information yet had distinct time courses. Our results suggest that, after interruption, lateralized visual representations of memoranda can stay active in

working memory for a short period without lateralized spatial attention before they are lost. These representations do not recover by the end of the trial and are presumed to be stored offline. By contrast, attention is directed away from the spatial location of memoranda immediately after the onset of the interruption but can recover later and may even contribute to the retrieval of information from offline storage. Thus, our results show that task-irrelevant interruption could motivate the transfer of information from active to passive storage. Moreover, the dissociation between CDA and lateralized alpha power further emphasizes that these neural markers distinctly contribute to the maintenance of information in working memory and may distinctly protect actively maintained memories from interruption.

Acknowledgments

Research was supported by National Institutes of Mental Health grant 5ROI MH087214-08 and Office of Naval Research grant N00014-12-1-0972. All authors conceived and designed experiments and drafted and revised the article. N. H. and T. F. W. performed analyses. N. H. collected the data.

Reprint requests should be sent to Nicole Hakim, University of Chicago, 940 East 57th Street, Chicago, IL 60637, or via e-mail: nhakim@uchicago.edu.

REFERENCES

Adam, K. C., Mance, I., Fukuda, K., & Vogel, E. K. (2015). The contribution of attentional lapses to individual differences in visual working memory capacity. *Journal of Cognitive Neuroscience*, *27*, 1601–1616.

Andrews, A., Ratwani, R., & Trafton, G. (2009). Recovering from interruptions: Does alert type matter? *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *53*, 409–413.

Bisley, J. W., & Goldberg, M. E. (2003). Neuronal activity in the lateral intraparietal area and spatial attention. *Science*, *299*, 81–86.

Bisley, J. W., Zaksas, D., Droll, J. A., & Pasternak, T. (2004). Activity of neurons in cortical area MT during a memory for motion task. *Journal of Neurophysiology*, *91*, 286–300.

Chun, M. M., & Turk-Browne, N. B. (2007). Interactions between attention and memory. *Current Opinion in Neurobiology*, *17*, 177–184.

Clapp, W. C., Rubens, M. T., & Gazzaley, A. (2010). Mechanisms of working memory disruption by external interference. *Cerebral Cortex*, *20*, 859–872.

Cowan, N. (2011). The focus of attention as observed in visual working memory tasks: Making sense of competing claims. *Neuropsychologia*, *49*, 1401–1406.

De Fockert, J. W., Rees, G., Frith, C. D., & Lavie, N. (2001). The role of working memory in visual selective attention. *Science*, *291*, 1803–1806.

Feldmann-Wüstefeld, T., & Schubö, A. (2013). Context homogeneity facilitates both distractor inhibition and target enhancement. *Journal of Vision*, *13*, 11.

Feldmann-Wüstefeld, T., & Vogel, E. K. (2018). Neural evidence for the contribution of active suppression during working memory filtering. *Cerebral Cortex*, *29*, 529–543.

Feldmann-Wüstefeld, T., Vogel, E. K., & Awh, E. (2018). Contralateral delay activity indexes working memory storage, not the current focus of spatial attention. *Journal of Cognitive Neuroscience*, *30*, 1185–1196.

Folk, C. L., & Remington, R. W. (2015). Unexpected abrupt onsets can override a top–down set for color. *Journal of Experimental Psychology: Human Perception and Performance*, *41*, 1153–1165.

Foster, J. J., Bsales, E. M., Jaffe, R. J., & Awh, E. (2017). Alpha-band activity reveals spontaneous representations of spatial position in visual working memory. *Current Biology*, *27*, 3216–3223.

Foster, J. J., Sutterer, D. W., Serences, J. T., Vogel, E. K., & Awh, E. (2016). The topography of alpha-band activity tracks the content of spatial working memory. *Journal of Neurophysiology*, *115*, 168–177.

Fukuda, K., Kang, M. S., & Woodman, G. F. (2016). Distinct neural mechanisms for spatially lateralized and spatially global visual working memory representations. *Journal of Neurophysiology*, *116*, 1715–1727.

Fukuda, K., Mance, I., & Vogel, E. K. (2015). Alpha power modulation and event-related slow wave provide dissociable correlates of visual working memory. *Journal of Neuroscience*, *35*, 14009–14016.

Fukuda, K., & Vogel, E. K. (2019). Visual short-term memory capacity predicts the "bandwidth" of visual long-term memory encoding. *Memory & Cognition*, *47*, 1481–1497.

Gaspar, J. M., & McDonald, J. J. (2014). Suppression of salient objects prevents distraction in visual search. *Journal of Neuroscience*, *34*, 5658–5666.

Hakim, N., Adam, K. C., Gunseli, E., Awh, E., & Vogel, E. K. (2019). Dissecting the neural focus of attention reveals distinct processes for spatial attention and object-based storage in visual working memory. *Psychological Science*, *30*, 526–540.

Horstmann, G. (2005). Attentional capture by an unannounced color singleton depends on expectation discrepancy. *Journal of Experimental Psychology: Human Perception and Performance*, *31*, 1039–1060.

Ikkai, A., McCollough, A. W., & Vogel, E. K. (2010). Contralateral delay activity provides a neural measure of the number of representations in visual working memory. *Journal of Neurophysiology*, *103*, 1963–1968.

Luria, R., Balaban, H., Awh, E., & Vogel, E. K. (2016). The contralateral delay activity as a neural measure of visual working memory. *Neuroscience & Biobehavioral Reviews*, *62*, 100–108.

Mallett, R., & Lewis-Peacock, J. A. (2018). Behavioral decoding of working memory items inside and outside the focus of attention. *Annals of the New York Academy of Sciences*, *1424*, 256–267.

Marini, F., Chelazzi, L., & Maravita, A. (2013). The costly filtering of potential distraction: Evidence for a supramodal mechanism. *Journal of Experimental Psychology: General*, *142*, 906–922.

Müller, H. J., Geyer, T., Zehetleitner, M., & Krummenacher, J. (2009). Attentional capture by salient color singleton distractors is modulated by top–down dimensional set. *Journal of Experimental Psychology: Human Perception and Performance*, *35*, 1–16.

Murray, A. M., Nobre, A. C., Clark, I. A., Cravo, A. M., & Stokes, M. G. (2013). Attention restores discrete items to visual short-term memory. *Psychological Science*, *24*, 550–556.

Olivers, C. N. (2008). Interactions between visual working memory and visual attention. *Frontiers in Biological Science*, *13*, 1182–1191.

Postle, B. R., D'Esposito, M., & Corkin, S. (2005). Effects of verbal and nonverbal interference on spatial and object visual working memory. *Memory and Cognition*, *33*, 203–212.

Sawaki, R., & Luck, S. J. (2012). Active suppression of distractors that match the contents of visual working memory. *Visual Cognition*, *19*, 1–14.

Thut, G., Nietzel, A., Brandt, S. A., & Pascual-Leone, A. (2006). Alpha band electroencephalographic activity over occipital cortex indexes visuospatial attention bias and predicts visual target detection. *Journal of Neuroscience*, *26*, 9494–9502.

Töllner, T., Müller, H. J., & Zehetleitner, M. (2012). Top–down dimensional weight set determines the capture of visual attention: Evidence from the PCN component. *Cerebral Cortex*, *22*, 1554–1563.

van Moorselaar, D., Foster, J. J., Sutterer, D. W., Theeuwes, J., Olivers, C. N. L., & Awh, E. (2018). Spatially selective alpha oscillations reveal moment-by-moment trade-offs between working memory and attention. *Journal of Cognitive Neuroscience*, *30*, 256–266.

Vogel, E. K., & Machizawa, M. G. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature*, *428*, 748–751.

Vogel, E. K., McCollough, A. W., & Machizawa, M. G. (2005). Neural measures reveal individual differences in controlling access to working memory. *Nature*, *438*, 500–503.

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2006). The time course of consolidation in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *32*, 1436–1451.

Williams, M., & Woodman, G. F. (2012). Directed forgetting and directed remembering in visual working memory. *Journal of Experimental Psychology: Learning Memory and Cognition*, *38*, 1206–1220.

Woodman, G. F., & Chun, M. M. (2006). The role of working memory and long-term memory in visual search. *Visual Cognition*, *14*, 808–830.

Worden, M. S., Foxe, J. J., Wang, N., & Simpson, G. V. (2000). Anticipatory biasing of visuospatial attention indexed by retinotopically specific alpha-band electroencephalography increases over occipital cortex. *Journal of Neuroscience*, *20*, RC63.