bioRxiv preprint doi: https://doi.org/10.1101/320952; this version posted May 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

# Unattended but actively stored: EEG dynamics reveal a dissociation between selective attention and storage in working memory

Gunseli, E.*, Fahrenfort, J., van Moorselaar, D., Daoultzis, K., Meeter, M., & Olivers, C. N. L.

* Corresponding Author:
Eren Gunseli
Columbia University
Gunseli.eren@gmail.com

Data is available on this link:
https://osf.io/bgpxc/?view_only=f4ab389a402d4138a3de5c1622a8e7dd

**Abstract**
Selective attention plays a prominent role in prioritizing information in working memory (WM), improving performance for attended representations. However, the consequences of selection for unattended WM representations are less clear, with mixed findings regarding information loss for unattended items. Here we tested the theory that within WM, selectively attending to an item and the decision to stop storing another item are independent mechanisms. We recorded EEG while participants performed a WM recall task in which the item most likely to be tested was cued retrospectively. By manipulating retro-cue reliability (i.e. the ratio of valid to invalid cue trials) we varied the incentive to retain uncued items. The cued item was initially attended equally following highly reliable and less reliable cues, as indexed by contralateral alpha (8-14 Hz) power suppression. Non-cued items were dropped from WM, as indexed by contralateral delay activity (CDA), but only for highly reliable cues. Later in the retention interval this pattern reversed. Selective attention was sustained only following highly reliably cues, while uncued items were dropped also for less reliable cues. These results show that attention and storage in WM are distinct processes that can behave differently depending on the relative importance of WM representations.

bioRxiv preprint doi: https://doi.org/10.1101/320952; this version posted May 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

## Introduction

Working memory (WM) is essential to storing and manipulating information online for a variety of cognitive tasks[1–4]. However, its capacity is limited[5,6] and thus only the most task-relevant information should be selected for storage in WM[7,8]. Attention is the mechanism by which task-relevant representations are prioritized and there is now a large body of evidence showing that attention and WM are heavily intertwined[9–12], such that attention may be crucial to successfully maintain an item in WM[13–20]. However, more recently alternative theoretical frameworks have been proposed that argue that storage of an item in WM should be dissociated from prioritization of (i.e. attending to) that item[21–26]. Thus, there is no consensus yet on the relationship between WM and attention.

Much of the evidence for a central role of selective attention in WM storage comes from studies using retrospective cues. Such "retro-cues" are presented after the to-be-remembered items have been taken away and indicate which of the memory representations is most likely to be tested and thus is the most task-relevant. Because retro-cues are presented after memory encoding, they act on stored WM representations rather than on encoding of stimuli. Nevertheless, retro-cues have been suggested to result in the attentional selection of the cued representation within WM in a similar way as attentional selection operates during perception, relying on highly overlapping neural networks[27–29]. This selection in turn has been claimed to improve storage and/or increase the accessibility of the cued item within WM[30–32]. The behavioral consequence is better memory performance for the attended representation compared to a 'no-cue' or neutral condition where all items are presumably equally attended[33–35].

The finding that retrospectively cueing attention to a representation improves memory performance does not in itself prove that attention plays a necessary role in the maintenance of that representation. For that, it is necessary to show that unattended items actually suffer from attention being cued elsewhere, relative to when attention is directed equally to all items. However, so far, the fate of *unattended* WM representations has been unclear. Memory performance for unattended representations can be tested by probing a non-cued representation on a minority of the trials. A lower memory performance on these *invalid* cue trials compared to *neutral* or no-cue trials is referred to as an 'invalid cueing cost'. Such invalid cueing costs have indeed been found in some studies, and have been taken as evidence that attention is necessary for WM storage[36–38]. However, using very similar cueing procedures, a number of other studies did not find such invalid cueing costs[21,23,30], suggesting a dissociation between storage and selection.

We recently proposed that the fate of non-cued items might depend on their perceived future relevance as inferred from the reliability of the retro-cue (i.e. the proportion of valid to invalid retro-cue trials)[40]. Typically, studies that failed to observe invalid cueing costs used lower retro-cue reliabilities[30,39,41] than studies that observed invalid cueing costs[31,36,38,42]. In a behavioral study, we observed invalid cueing costs only when the retro-cue had a high reliability (i.e., 80%

bioRxiv preprint doi: https://doi.org/10.1101/320952; this version posted May 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

valid), but not when it had a lower, but still above-chance reliability (i.e., 50% valid, with chance level being at 25% in both conditions). While the presence of invalid retro-cue costs varied with retro-cue reliability, benefits of *valid* retro-cues were present in both conditions, though they were larger for 80% valid cues. This can explain the discrepant findings in the literature if we assume that attending to an item in WM can be dissociated from the decision to either continue or cease storage of remaining items. For both moderately and highly reliable cues it is beneficial to attend to an item, as it is more likely to be tested than uncued items. However, only for highly reliable cues it is also worth dropping the uncued items from memory, while for moderately reliable cues it is actually worth holding on to the uncued items.

Although our behavioral work provides initial evidence for the idea that attention to and storage of an item should be dissociated when interpreting the effects of retro-cueing, there is an alternative scenario that can explain the reliability effects on performance for uncued items, in which increasing the retro-cue reliability results in more attention to the selected representation without affecting the probability with which the unattended representations are dropped from WM. Under this scenario, items in principle remain stored in WM regardless of cue reliability, but they become more vulnerable to interference from the test display when unattended. The test display is in itself a stimulus that may overwrite a fragile memory representation, and it has been proposed that attention protects against such interference[43–46]. This would then result in larger invalid cueing costs for highly reliable cues, even if unattended items were still stored until the test display. While a differential storage account predicts that the decision to drop an uncued item is made during the retention interval, the protection against interference account predicts that nothing happens to uncued items during the retention interval and that performance differences result from processes during test. Because behavioral methods only measure the final outcome, they are blind to the underlying mechanisms during retention and therefore cannot differentiate between these scenarios.

To more directly investigate if and how retro-cue reliability affects attention and storage in WM prior to the test, we used EEG recordings to measure these processes in a time-resolved manner during the retention interval. The experimental procedure is shown in Figure 1. We used a continuous report WM task to obtain a sensitive measure of memory performance. This also enabled us to model the error distribution and estimate the probability of storing the tested WM representation and its fidelity[5,47]. The memory display contained three line segments of different orientations, one on the vertical midline and the other two presented left and right from fixation. After a blank interval, a retro-cue indicated which of the memory representations was most likely to be tested by retrospectively pointing to its location in the memory display. Only lateral cue trials were included for the EEG analysis since both of our EEG indices of interest (see below) required a lateral asymmetry in the location of the attended and stored item. Critically, to vary the incentive to also retain the uncued items, we manipulated the retro-cue reliability (i.e. the proportion of valid to invalid trials) across blocks: The cue was 50% valid in half the number of blocks, and 80% valid in the other half.

bioRxiv preprint doi: https://doi.org/10.1101/320952; this version posted May 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

As a proxy for *attention* being directed within WM we used contralateral power suppression in the alpha band (8-14 Hz). Alpha power over the parietal-occipital electrodes on the hemisphere contralateral to the attended item has been found to be more negative relative to the ipsilateral electrodes, both during perception and during post-perception within WM[48–53]. We hypothesized that if the cued item is attended more during storage for highly reliable cues, then we should observe a larger contralateral alpha suppression for highly reliable retro-cues. As a marker for *storage* we focused on the CDA, which is a sustained negativity over the parietal-occipital electrodes on the hemisphere contralateral to remembered stimuli. It has been observed to be sensitive to the visual WM load, and converging evidence suggests that it is an index of visual WM storage[54–56]. We reasoned that if non-cued representations are dropped following a retro-cue, then a CDA should emerge, since dropping an item on one side results in an imbalance in the number of items stored in each hemisphere[57]. If, as we hypothesized, the likelihood of dropping an item depends on retro-cue reliability, we should see a CDA emerge in the high reliability condition, but not in the low-reliability condition. Alternatively, if retro-cue reliability has no effect on storage, we should see no differential CDA, and only find attentional effects as expressed through alpha suppression.

## Method

Thirty-two healthy volunteers participated in the experiment for course credit or monetary compensation. Two participants were excluded; one due to excessive noise in their EEG recordings and one due to poor behavioral performance (see Data Analyses), leaving 30 participants of whom the data was analyzed. The study was conducted in accordance with the Declaration of Helsinki and was approved by the faculty's Ethical Committee. Written informed consent was obtained prior to the experiment. Data sets are available online on Open Science Framework at https://osf.io/bgpxc/?view_only=3b8dd8f9e4fa42d68ac84db90f76e25d

The procedure is shown in Figure 1. Each trial started with the presentation of the fixation circle of radius .33º, for a duration jittered between 1200-1600 ms. Then, the memory display was presented for 350 ms. It consisted of three black oriented bars (2.08º x 0.25º visual angle) located at 60 (top right), 180 (bottom) and 300 (top left) degrees relative to the top of an imaginary circle of radius 3.50º. We used a memory load of three items in order to tax WM without contaminating measurements with non-encoded items. The orientation of each bar was chosen at random with the restriction that bars within the same trial differed by at least 10º. The retro-cue was presented for 100 ms following a blank interval of 650 ms during which only the fixation circle was presented. The retro-cue was identical to the fixation circle except that one quarter (90º) was now filled with either red, 27.08 Cd/m$^2$, or green, 24.10 Cd/m$^2$, depending on the reliability condition (order counterbalanced). For the initial practice phase where the cue was 100% valid, the retro-cue fill color was orange (53.46 Cd/m$^2$). Following the retro-cue, there was a blank interval of 900 ms in which only the fixation circle was presented. Then the test display

bioRxiv preprint doi: https://doi.org/10.1101/320952; this version posted May 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

was presented till response. It contained a probe cue pointing to the location of the tested representation and a randomly oriented probe bar that were both presented at the center of the screen. This probe cue was the same as the retro-cue except that the filling color was white. Participants were asked to indicate the orientation of the bar at the tested location as precise as possible by rotating the probe bar using the mouse and pressing the left mouse button. After a mouse response was made, the correct orientation was indicated by a central white bar for 100 ms. The screen was empty during the inter-trial interval which was jittered between 1200-1600 ms.

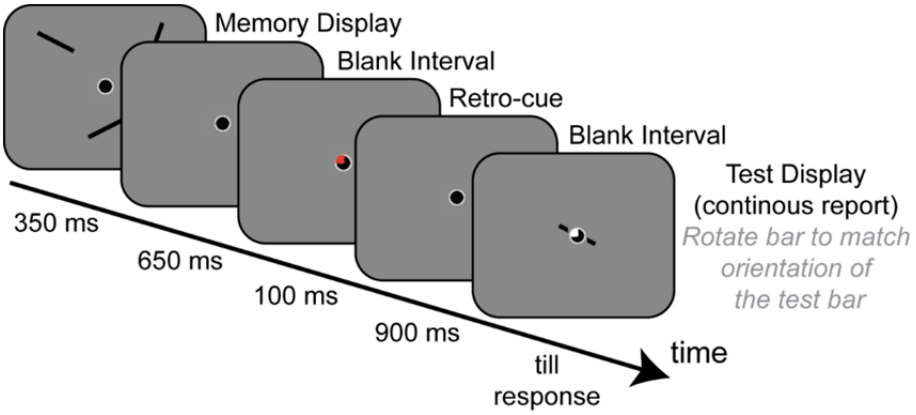


**Figure 1**. The retro-cue experimental procedure. Participants were asked to remember the three orientations shown in the memory display. After a blank interval, a retro-cue was presented pointing to the location of the item (in this example top-left) that was most likely to be tested. Retro-cues were not always valid. Following a second blank interval the test display was presented during which participants were asked to rotate a randomly-oriented bar to match the orientation of the tested item (which in this example is the item presented on top-left, hence the retro-cue was valid).

The retro-cue was 80% valid for half of the experiment and 50% valid for the other half (order counterbalanced). There were 10 blocks of 50 trials. Each validity condition (i.e. valid and invalid) was randomly intermixed within each block. Before each reliability condition, participants were informed about the validity ratio of the retro-cue and they performed a practice session of 25 trials to get used to this particular validity ratio. Moreover, at the beginning of the experiment, there was an initial practice session of 25 trials with a 100% valid cue to make participants familiar with using the experimental procedure. At the end of each block, participants received feedback on block average and grand average error (i.e. the difference between the original tested orientation and the responded orientation).

## EEG Data Acquisition

The electroencephalogram (EEG) and electro-oculogram (EOG) were recorded from 70 sintered –AG/AgCl electrodes positioned at 64 standard International 10/20 System sites and 6 external locations mentioned below, using the Biosemi ActiveTwo system (Biosemi, Amsterdam, the

bioRxiv preprint doi: https://doi.org/10.1101/320952; this version posted May 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Netherlands). No impedance measurements or gain adjustments are needed with the ActiveTwo system (www.biosemi.com). The vertical EOG (VEOG) was recorded from electrodes located 2 cm above and below the right eye, and the horizontal EOG (HEOG) was recorded from electrodes 1 cm lateral to the external canthi. The VEOG was used in the detection of blink artifacts, and the HEOG was used in the detection of horizontal eye movement artifacts. Electrophysiological signals were digitized at 512 Hz.

## Data Analysis

### *Behavior*

Error scores on the memory test were calculated as the difference between the original orientation of the tested memory bar and the orientation of the response. One participant with an average absolute error value higher than 2.5 standard deviation above the grand average of the group was excluded from analysis. The error scores were entered into the swap model[58], which assumes that there are three sources of error in a continuous report memory recall task: (1) a Gaussian variability in memory for the tested item, reflected in a circular analogue of the Gaussian distribution (i.e. von Mises distribution) centered around the tested orientation, (2) Gaussian variability in memory for the mistakenly reported non-tested items, which is reflected in a von Mises distribution centered around each non-tested orientation, and (3) a flat distribution due to random guessing. Using MemToolbox (memtoolbox.org)[59] we estimated the guess rate (i.e. reporting a random orientation), swap rate (i.e. reporting a non-tested representation), and sd (inverse of precision) that is estimated by the width of the response distribution around the target. Absolute raw error, guess rate, swap rate and sd estimates for the tested item were entered into four separate repeated-measures ANOVAs with the within-subjects factors of retro-cue reliability (80% valid; 50% valid) and retro-cue validity (valid; invalid). Moreover, the validity effect on each model estimate was tested by comparing valid and invalid trial estimates using paired-samples t-tests.

### *EEG analysis: General*

Only lateral cue trials were included for the EEG analysis since both of our EEG indices of interest require a lateral asymmetry in the location of the attended or stored item. All EEG analyses were carried out using the EEGLAB toolbox[60] and custom scripts implemented in MATLAB (The MathWorks, Inc., Natick, MA). Due to unknown reasons, there were three subjects who had parts of EEG data missing (10, 11 and 26 trials). Noisy electrodes were interpolated using the "eeg_interp.m" function of EEGLAB with the spherical interpolation method, which resulted in the interpolation of three electrodes each for two subjects (FC2, C6, PO3; CP3, PO3, P4). None of these electrodes were used in the statistical analysis. EEG waveforms at the electrodes of interest (P7/8, PO7/8, and O1/2) were visually inspected for recording artifacts (muscle noise and slow drifts) and EOG waveforms were visually inspected for ocular artifacts (blinks and eye movements) in the absence of any knowledge about the

bioRxiv preprint doi: https://doi.org/10.1101/320952; this version posted May 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

conditions. Individuals were excluded from analyses if, after all the artifact rejections, the remaining number of trials per condition was lower than 80 trials. This led to the rejection of one participant. For the remaining participants, on average 9.8% of all trials were rejected due to artifacts, leaving on average 139 and 141 lateral cue trials for analysis (with a minimum of 105 and 111 trials), for 80% valid and 50% valid blocks respectively.

*ERP analysis: CDA*

ERPs were computed with respect to a 200 ms pre-stimulus baseline period, between -500 to 1500 ms around the retro-cue display and were re-referenced offline to the average of left and right mastoids. The data was filtered with an IIR Butterworth filter with a bandpass of .01 – 40 Hz. Signal was resampled at 500 Hz using "pop_resample.m" function of EEGLAB.

The CDA was calculated as the difference waves between electrode sites contralateral versus ipsilateral to the location of the retro-cued item. Previous studies measuring the CDA have typically found maximal values at posterior/occipital electrodes and started measuring the CDA at ~300-400 ms from the onset of the memory display following the N2pc (~200-300 ms) which signals individuation of selected items[54,61,62]. Based on these studies and visual inspection of the topographic distribution of lateralized voltage in posterior/occipital regions we calculated the CDA at P7/8, PO7/8, and O1/2 as contralateral minus ipsilateral to the retro-cud location starting from 400 ms following the retro-cue onset till the onset of the test display (i.e. 900 ms after cue onset). The CDA averaged across electrode pairs and times of interest was entered into a repeated-measures ANOVA with the within-subjects factor of retro-cue reliability (80% valid; 50% valid). Average CDA values were also tested against zero using one-sample t-tests. Additionally, given that CDA is a sustained negativity at electrodes contralateral to items stored in WM, we hypothesized that if the reliability effect on CDA reflects a boost for the representation of the cued item, then it should be evident in the signal contralateral to the *cued* item, while if it reflects dropping of the non-cued item then the reliability effect should be observed in the signal contralateral to the *non-cued* item (i.e. ipsilateral to the cued item). To test this, the average contralateral and ipsilateral signals were entered into a repeated-measures ANOVA with within-subjects factors of laterality (contralateral; ipsilateral to the cued item) and retro-cue reliability (80% valid; 50% valid).

In order to investigate the dynamic time course of the reliability effect, the CDA at each time point for each reliability condition were tested against chance and also against each other at a group level using cluster-based permutation testing by estimating the permutation *p*-value using a Monte Carlo randomization procedure[63]. For this analysis, we randomly shuffled the condition labels (e.g. 50% valid vs. 80% valid) 1000 times to approximate the null distribution of the t statistic. The *p*-value was the proportion of iterations out of 1000 where the absolute randomly shuffled condition difference was larger than the absolute actual condition difference (two-sided). Multiple comparisons correction was established using cluster-based permutation testing. First, four or more temporally adjacent data points with a p-value smaller than 0.05 were

bioRxiv preprint doi: https://doi.org/10.1101/320952; this version posted May 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

clustered together. Then, a cluster-level statistic was calculated by taking the sum of the t-values within each cluster, separately for positive and negative clusters. The p-value for each cluster was calculated as the number of times the sum of the absolute t-values within the cluster under random permutation exceeds that of the t-values within the observed cluster. A cluster was considered significant if the calculated p-value was smaller than 0.05.

*Time-frequency power in the alpha band (8-14 Hz)*

Time-frequency analysis was performed using the same trials as in the CDA analysis. Prior to the calculation of time-frequency representations (TFRs), the signal was epoched between -1000 to 2000 ms around the onset of the cue display. We chose a larger window compared to the CDA analysis in order to avoid contaminating the results from edge artifacts that result from applying a band-pass filter at the edges of an epoch [60]. To isolate alpha-band activity, we bandpass-filtered the raw EEG between 8 and 14 Hz using "eegfilt.m" (EEGLAB Toolbox)[60]. This function filters the data using a two-way least-squares FIR filtering. Then, in order to produce a complex analytic signal, we applied a Hilbert transform (MATLAB Signal Processing Toolbox) to the band-pass filtered data. We computed instantaneous power by taking square of the complex magnitude of the complex analytic signal. After calculating the power, the epochs were reduced to -500 to 1500 around the retro-cue display. Power data was baseline normalized separately for each condition (i.e. 50% valid left-cue, 50% valid right-cue; 80% valid left-cue; 80% valid right-cue) with decibel (dB) conversion, using -400 to -100 ms relative to the retro-cue onset as baseline[64]. The dB normalized data was averaged separately for contralateral and ipsilateral in respect to the side of the cued item at the electrode pairs of interest (P7/P8, PO7/PO8, O1/O2). Contralateral alpha suppression was calculated as the difference between the contralateral and ipsilateral dB normalized power values.

Contralateral alpha band power averaged across electrodes and times of interest (400 – 900 ms, which is chosen to be the same time interval as the CDA analysis) were compared between 50% and 80% valid conditions using a repeated measures ANOVA. Average contralateral alpha band power values were also tested against zero using one-sample t-tests. Since we performed a reliability analysis separately for contralateral and ipsilateral signal for the CDA, for completeness we also performed it for lateral alpha power. Power values contralateral and ipsilateral power relative to the direction of the cued item were averaged across the time window of interest entered into a repeated-measures ANOVA with within-subjects factors of laterality (contralateral; ipsilateral to the cued item) and retro-cue reliability (80% valid; 50% valid). Lastly, contralateral power suppression at each time point for each reliability condition were tested against chance and against each other at a group level with the same cluster-based permutation test as in the CDA analysis.

bioRxiv preprint doi: https://doi.org/10.1101/320952; this version posted May 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

*Correlation between average CDA and average contralateral alpha suppression*

In order to test whether selective attention within WM predicts storage in WM on a subject level, we performed Pearson correlations between the average CDA and the average contralateral alpha suppression across subjects, separately for 50% valid and 80% valid blocks.

*Comparing time courses of the reliability effect on CDA and contralateral alpha suppression*

Upon observing visually different patterns of reliability effect on the CDA and the contralateral alpha suppression across time, we performed a post-hoc analysis to test this difference statistically. First, we split the time window of analysis into early (400-600 ms) vs. late (700-900ms). We used a 100 ms buffer between early and late time windows, but the results were the same when the retention interval was split into two without any buffer (i.e. 400-650 and 650-900 ms). The average CDA and contralateral alpha values across time windows were entered into a repeated-measures ANOVA with within-subjects factors of EEG index (CDA; contralateral alpha suppression), time window (early; late), and retro-cue reliability (50% valid; 80% valid).

## Results

### Behavior

*Raw Error*

Figure 2A shows the distribution of errors (i.e. the difference between the original tested orientation and the response) for each condition using bins of 15 degrees of deviations. There was a main effect of validity on error, $F(1, 29) = 27.75$, $p < .001$, $\eta_p^2 = .49$. Errors were larger on invalid compared to valid cue trials. Importantly, this validity effect (i.e. error on invalid trials minus the error on valid trials) was larger when the cue was 80% valid compared to when it was 50% valid, as indicated by a validity x reliability interaction, $F(1, 29) = 6.49$, $p = .016$, $\eta_p^2 = .18$. There was no main effect of reliability on error, $F(1, 29) = 2.41$, $p = .131$, $\eta_p^2 = .07$. In sum, the effect of retro-cues on raw error was larger for cues that were more reliable.

*Guess rate*

The average probability of guessing (i.e. responding randomly) in each condition is shown in Figure 2B. Guess rate was larger in invalid cue trials than in valid cue trials, $F(1, 29) = 6.35$, $p = .017$, $\eta_p^2 = .18$. Guess rate was marginally larger on 80% valid blocks than on 50% valid blocks, $F(1, 29) = 3.78$, $p = .0.62$, $\eta_p^2 = .12$. There was a validity x reliability interaction, $F(1, 29) = 6.11$, $p = .020$, $\eta_p^2 = .17$. The invalid cueing cost on guess rate (i.e. the guess rate on invalid trials minus the guess rate on valid trials) was larger in 80% valid blocks than in 50% valid blocks, $t(29) = 2.48$, $p = .019$. The invalid cueing cost was significant only in 80% valid blocks, $t(29) = 2.52$, $p$

bioRxiv preprint doi: https://doi.org/10.1101/320952; this version posted May 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

=.017, but not in 50% valid blocks, $t(29) = 0.26$, $p = .794$. Compared to 50% valid blocks, in 80% valid blocks the guess rate was smaller in valid trials, $t(29) = 2.33$, $p = .027$, and larger in invalid trials, $t(29) = 2.22$, $p = .035$. In sum, the invalid cueing cost on guess rate was exclusive to highly reliable cues.

*Swap rate*

The average swap rate (i.e. probability of reporting a non-tested item instead of the tested item) in each condition is shown in Figure 2C. Swap rate was larger on invalid compared to valid cue trials, $F(1, 29) = 15.67$, $p < .001$, $\eta_p^2 = 0.35$. Swap rate was marginally larger on 80% valid blocks than on 50% valid blocks, $F(1, 29) = 3.28$, $p = 0.081$, $\eta_p^2 = .10$. There was a marginally significant validity x reliability interaction, $F(1, 29) = 3.65$, $p = 0.066$, $\eta_p^2 = .12$ reflecting that the invalid cueing cost on swap rate (i.e. the swap rate on invalid trials minus the swap rate on valid trials), although present in both reliability conditions ($p$s < 0.003), was larger in 80% valid blocks than in 50% valid blocks. In sum, there was a trend for a larger effect of invalid cueing on the probability of reporting a non-tested item for retro-cues that were more reliable.

*Precision*

The average sd (i.e. inverse of precision) in each condition is shown in Figure 2D. The sd was larger (i.e. precision was worse) in invalid compared to valid cue trials, $F(1, 29) = 8.41$, $p = .007$, $\eta_p^2 = .22$. There was no effect of reliability, $F(1, 29) = .02$, $p = .89$, $\eta_p^2 < .01$, nor a validity x reliability interaction, $F(1, 29) = .4$, $p = .53$, $\eta_p^2 = .01$.

bioRxiv preprint doi: https://doi.org/10.1101/320952; this version posted May 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

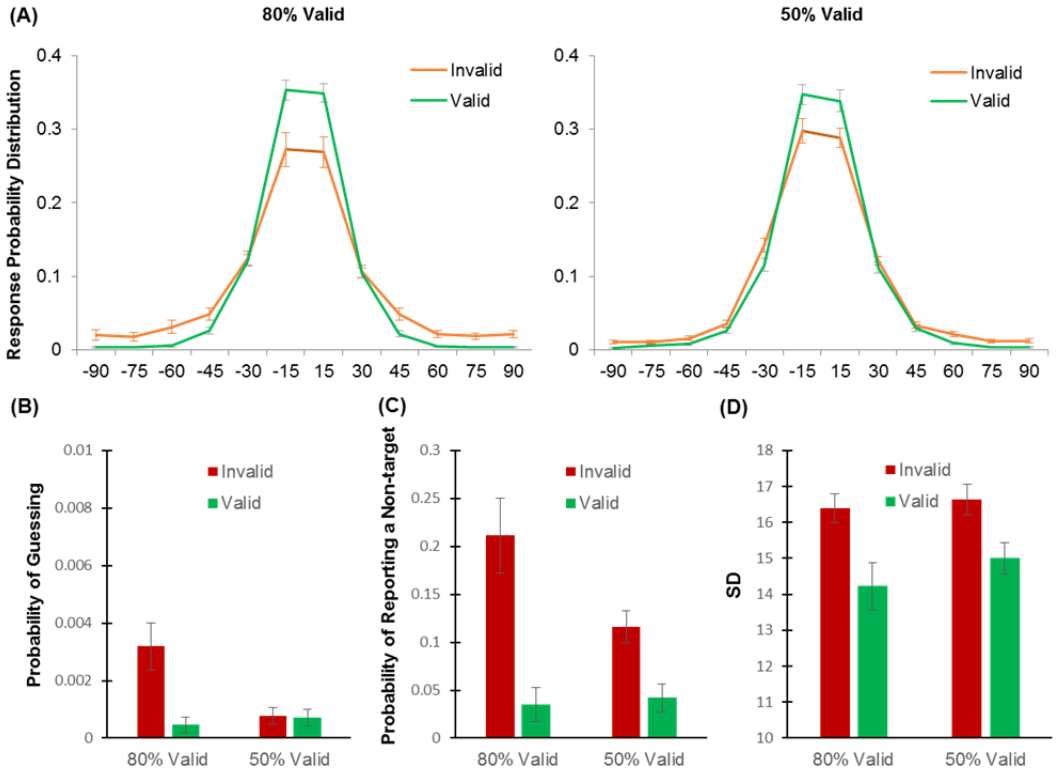


**Figure 2. (A)** Distribution of errors relative to the target (i.e. tested) orientation for the 80% valid (left panel) and 50% valid (right panel) conditions**. (B)** Probability of randomly responding (i.e. guess rate), **(C)** probability of reporting a non-target (i.e. swap rate), and **(D)** SD (i.e. inverse of precision) estimates for the target in each condition. The invalid and valid trials are shown in different colors. The error bars represent standard errors of the mean for normalized data, i.e. corrected for between-subjects variance (Cousineau, 2005). Retro-cue validity effect was larger for highly reliable cues than less reliable cues both for guess rate and swap rate.

## Electrophysiology

*Selective attention was allocated to the cued item independent of retro-cue reliability, but was sustained only for highly reliable cues*

Figure 3A shows the contralateral alpha band (8-14 Hz) suppression (i.e. the difference between the contralateral and ipsilateral power) with respect to the position of the cued item averaged across the electrode pairs of interest (P7/8, PO7/8 and O1/2). The permutation tests showed significant contralateral alpha suppression in both the 80% valid condition (significant time points: 400-850 ms) and the 50% valid condition (significant time points: 438-584 ms), with initially no difference between the validity conditions. After about 600 ms however, the contralateral alpha suppression in 50% valid condition dropped back to the baseline, resulting in a significant difference between the 50% valid and 80% valid conditions (significant time points: 702-900 ms). As a consequence, when the mean contralateral alpha suppression was averaged

bioRxiv preprint doi: https://doi.org/10.1101/320952; this version posted May 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

across the entire time window of interest (400-900 ms), it was marginally stronger (i.e. more negative) in 80% valid blocks (*M*= -0.22, *SD* = 0.42) than in 50% valid blocks (*M*= -0.05, *SD* = 0.39), $F(1, 29) = 3.59$, $p = .068$, $\eta_p^2 = .11$. Moreover, as seen in Figure 4 the retro-cue reliability affected the signal on the contralateral hemisphere relative to the cued item, $t(29) = 2.18$, $p = 0.037$, but not the ipsilateral hemisphere, $t(29) = 1.16$, $p = 0.25$. These results suggest that early in the trial the cued item was attended independent of retro-cue reliability, but that attentional prioritization was sustained for a longer period of time when cue was more reliable.

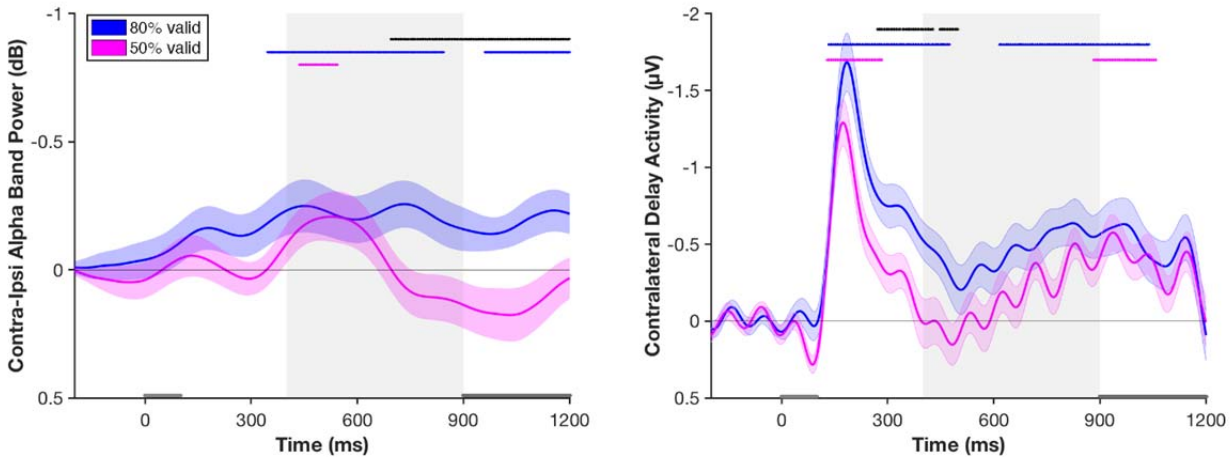


**Figure 3. (A)** Contralateral alpha suppression, and **(B)** CDA as indices of selective attention and storage in WM respectively, both time-locked to the onset of the retro-cue, are shown in different colors for 80% valid and 50% valid conditions. The gray area shows the time window of interest (400-900 ms). The gray rectangles on the x-axis show the timing of the retro-cue (0-100 ms) and the test display (from 900 ms till response, which extends till 1200 ms on the plots). The CDA was low-pass filtered at 12 Hz for plotting purposes. Markers along the top of each plot indicate the time points at which either the difference between the EEG measures in 80% valid and 50 % valid conditions (black) or the EEG measure itself for each condition (blue for 80% valid and magenta for 50% valid) were significantly different than zero as determined by a cluster-based permutation test ($p<0.05$; two-tailed). For highly reliable cues, the cued item was attended and non-cued items were dropped from WM. For less reliable cues, non-cued items were initially unattended but were kept in WM until about the onset of the test display.

*Non-cued items were dropped from WM earlier and more often after highly reliable cues*

Figure 3B shows the difference between contralateral and ipsilateral waveforms in respect to the location of the retro-cued item, averaged across the electrode pairs of interest (P7/8, PO7/8 and O1/2). First, following the traditional approach we averaged the CDA across the whole analysis time window. There was a significantly larger CDA (i.e. more negative) in 80% valid blocks than in 50% valid blocks, $F(1, 29) = 6.29$, $p = 0.018$, $\eta_p^2 = 0.18$. In fact, there was a significant CDA only in the 80% valid condition (*M*= -0.45, *SD* = 0.61), $t(29) = -4.04$, $p < 0.001$, and not in the 50% condition (*M*= -0.15, *SD* = 0.54), $t(29) = -1.58$, $p = 0.126$. Furthermore, as seen in

bioRxiv preprint doi: https://doi.org/10.1101/320952; this version posted May 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 4, the retro-cue reliability affected the signal on the ipsilateral hemisphere relative to the cued item, $t(29) = 2.63$, $p = 0.013$, but not the contralateral hemisphere, $t(29) = 1.56$, $p = 0.13$, which is in contrast with the findings in typical CDA studies where participants store items presented in a single hemifield and the memory load mainly affects the signal contralateral to the memory items[65]. This dissociation in line with our conclusion that the CDA reflects dropping of the non-cued item instead of an attentional boost for the cued item. For a time resolved approach, we used cluster-based permutation test. This analysis showed that for 80% valid blocks there was a significant CDA both early and late during the CDA time interval (significant time points: 400-476 ms and 616-900 ms after the retro-cue onset), while for 50% valid blocks there was a CDA only later in the trial, moments before the test display (significant time points: 882-900 ms). Importantly, in contrast to the alpha suppression here the difference between 80% valid and 50% valid conditions was mainly significant early in the trial (400-430 ms and 446-500 ms) and not late in the trial. In sum, the CDA results suggest that non-cued items were dropped from WM right after the retro-cue after highly reliable cues and later in the retention interval after less reliable cues.

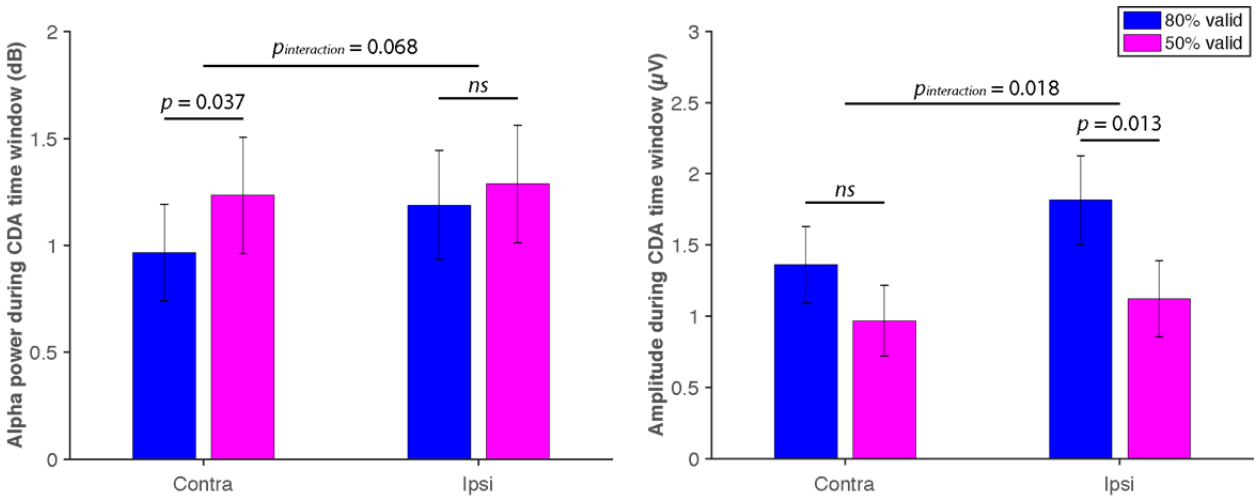


**Figure 4.** Lateral alpha power (left panel) and EEG waves (right panel) contralateral and ipsilateral to the side of the cued item averaged across the time window and electrodes of interest. 80% valid and 50% valid conditions are shown in different colors. Reliability effect was reflected in differences in signal at electrodes contralateral to the *non-cued* item for CDA, but contralateral to the *cued* item for contralateral alpha suppression. This is in line with the claim that the CDA reflects dropping of the non-cued item instead of an attentional boost for the cued item and that CDA and contralateral alpha suppression reflect different cognitive processes.

bioRxiv preprint doi: https://doi.org/10.1101/320952; this version posted May 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

*Comparing the time course of the reliability effect on CDA and Contralateral Alpha Suppression*

Permutation test results suggested that the reliability effect on the CDA emerged early in the retention interval and disappeared later in the trial, while the reliability effect on contralateral alpha suppression showed the opposite pattern: No difference early in the retention period, but a difference emerging later in the delay. In order to quantify the change of the retro-cue reliability effect across time and to compare it for the CDA and the contralateral alpha suppression we performed a repeated measured ANOVA with within-subjects factors of EEG index (CDA; contralateral alpha suppression), time window (early; late), and retro-cue reliability (50% valid; 80% valid) which revealed a significant 3-way interaction, $F(1, 29) = 7.02$, $p = .013$, $\eta_p^2 = .19$ (same results were obtained when early and late time windows were respectively defined as 400-650 ms and 650-900 ms; $F(1, 29) = 5.21$, $p = .03$, $\eta_p^2 = .15$). As shown in Figure 5, the reliability effect showed different patterns across CDA and contralateral alpha suppression, as it increased across time for contralateral alpha suppression while it decreased across time for CDA.

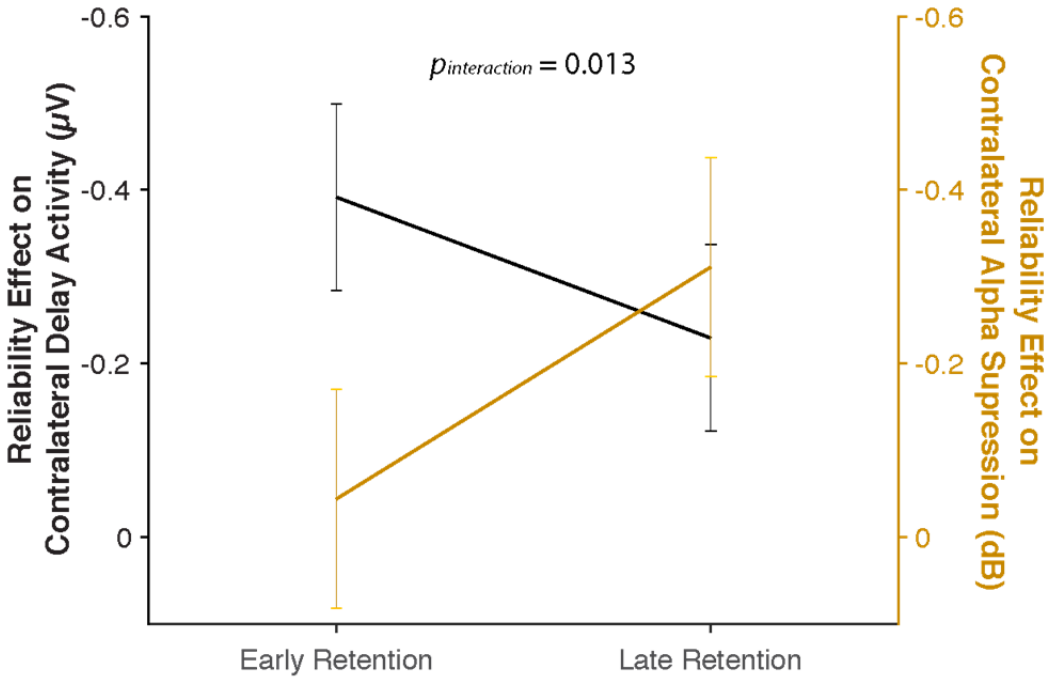


**Figure 5.** The retro-cue reliability effect (i.e., the difference between 80% valid and 50% valid conditions) on the CDA and the contralateral alpha suppression separately for early retention interval (400-600 ms) and late retention interval (700-900 ms). The reliability effect decreased over time for CDA while it increased over time for contralateral alpha suppression ($p$=0.013). Thus, CDA and contralateral alpha suppression can behave distinctly, in this case opposite, across time.

bioRxiv preprint doi: https://doi.org/10.1101/320952; this version posted May 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

*Correlations between CDA and Contralateral Alpha Suppression*

To further test whether selective attention within WM relates to storage, we ran across-subjects correlation analyses, correlating the average CDA and the average contralateral alpha suppression, separately for 80% valid and 50% valid blocks. There was no significant correlation for 80% valid blocks ($R^2 = 0.01$, $p = 0.54$) or 50% valid blocks ($R^2 = 0.12$, $p = 0.86$). Separate correlation analyses for early and late time windows also failed to reveal any correlation between the CDA and the contralateral alpha suppression ($R^2 < 0.03$, $ps > 0.342$). These results suggest that selective attention (as indexed by alpha suppression) did not predict storage (as indexed by the CDA) in WM on a participant level.

## Discussion

Selective attention has been claimed to be essential for WM storage[12,14,16,19,66,67]. Yet, results regarding the costs of allocating attention away from WM representations have been conflicting[30–32,36,37,39]. To shed new light on these discrepant results, we used EEG indices of spatial selective attention (i.e. contralateral alpha suppression) and storage (i.e. CDA) within WM when the most task-relevant representation was cued. Importantly, following our previous work that shows retro-cue costs are sensitive to the reliability of the cue, we manipulated the proportion of valid to invalid trials of the retro-cues across blocks (80% valid vs. 50% valid)[40]. We replicated our previous behavioral findings by showing that the retro-cue effect on recall probability of the uncued item depends on the reliability of the cue.

Here we show that these behavioral findings have a correlate in the amount of attention paid to the cued item on the one hand, and the dropping of non-cued items on the other. Specifically, right after the cue, contralateral alpha suppression was equal regardless of cue reliability, suggesting that the cue was used to direct attention to the cued representation as it presented useful information irrespective of whether the cue was 50% or 80% valid. However, contralateral alpha suppression then persisted throughout the retention period for highly reliable cues, while it dropped to baseline for less reliable cues. This indicates that attention was sustained on the cued item when participants could be reasonably sure that it would also be the tested item. As a measure of storage we took the CDA, which emerged early in the retention interval after highly reliable cues, consistent with a rapid drop of the uncued item. Later in retention, a CDA also emerged for low-reliability retro-cue condition, suggesting that eventually uncued items were also dropped after less reliable cues. Thus, the time-resolved nature of our EEG measures reveals that the reliability of the retro-cue had dissociable effects on the CDA and contralateral alpha suppression. Early in the retention interval, the CDA was stronger for highly reliable cues while contralateral alpha suppression was equal across reliability conditions. Later in the retention interval, contralateral alpha suppression was stronger for highly reliable cues, while the CDA showed no difference. Therefore, for highly reliable cues non-cued items were both unattended and dropped from WM. However, for less reliable cues non-cued items were initially unattended but kept in WM, while later they were lost from WM although now being attended. These results

bioRxiv preprint doi: https://doi.org/10.1101/320952; this version posted May 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

suggest that attentional selection of an item in WM is not accompanied by the loss of unattended items when there is a relatively high chance that these items could be relevant in the future.

Uncued items were unattended but kept in WM following less reliable retro-cues. To our knowledge, ours is the first study to observe simultaneous neural evidence for prioritization of an attended item and active storage of unattended items in WM. This finding is in contrast with the claims that suggest WM storage is a direct reflection of selective attention in WM[14,16,19,66,67], and supports the view that attentional prioritization of an item is a separate decision than dropping the remaining items[21–26]. This discrepancy between the two bodies of evidence is likely due to differences between the perceived future relevance of unattended items across studies, as studies that observed costs for unattended items used highly reliable cues, for which here we show that unattended items were dropped immediately. Yet, also for less reliable cues non-cued items were eventually dropped from WM right before the onset of the test display. There are two explanations for the delayed loss of unattended items for less reliable cues compared to highly reliable cues. First, non-cued items might have become more vulnerable to interference from other items in WM due to being initially unattended following the retro-cue. This, in turn might have resulted in the deterioration of non-cued items through the retention interval. This scenario is in line with the evidence that proposes selective attention protects WM items against inter-item interference during storage[24]. Alternatively, non-cued items might have been stored in WM for a longer duration in an attempt to create passive memory traces for these items. Passive memory traces have been claimed to be established over short durations for currently less relevant representations[68–70]. Later, with the anticipation of the test display they might have been deliberately dropped from WM in order to allocate all mnemonic resources to the most relevant item to protect it against perceptual interference by the test display. This strategy would be effective given previous findings that show smaller perceptual interference for smaller memory loads[43–45]. Importantly, although both of these explanations support the protective role of selective attention for storage in WM, they are not against our conclusion that selective attention and storage in WM are distinct constructs, as here we show that an item can be unattended but actively stored in WM at a given time.

Our results provide indirect evidence for the protection account that suggests selective attention protects WM items against perceptual interference[43–45]. The CDA was equal in 50% valid and 80% valid blocks at the end of the trial suggesting that by the onset of the test display uncued items were dropped from WM equally often across reliability conditions. However, behavioral measures suggested that non-cued items were forgotten more often and cued items were forgotten less often in 80% valid blocks than in 50% valid blocks. This discrepancy between the EEG and behavioral results can be explained by the different attentional prioritization of the cued item at the end of the trial across reliability conditions. Selective attention to the cued item was sustained only in 80% valid blocks. Thus, we suggest that allocating attention away from non-cued items made them more vulnerable to interference from the test display, resulting in lower

bioRxiv preprint doi: https://doi.org/10.1101/320952; this version posted May 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

behavioral performance when invalidly tested even though they were equally likely to be stored in WM prior to the test display.

Selective attention to the cued item was sustained till the end of the trial for highly reliable cues, but not for less reliable cues. We propose that the allocation of attention back to non-cued items for less reliable cues reflects an attempt to revive previously unattended items that were being lost. This is consistent with recent evidence that suggests weakly encoded representations, which are presumably also the weakly represented ones, are prioritized during WM retention in an attempt to prevent their loss[71]. Attentional reallocation to non-cued items was not observed for highly reliable cues. Given existing evidence that suggests the use of retro-cues is at least partly under strategic control[40,72,73], we claim that an item that is being lost is attended only when it is highly relevant for the ongoing task. Thus, our results provide evidence for the flexible nature of WM by showing that selective attention can be strategically adjusted based on the perceived future relevance of WM items. In addition to its protective function, selective attention in WM has also been suggested to increase the accessibility of the attended item in a way that it effectively guides behavior in the external world[22,23,74,75]. Thus, we argue that the presence of an additional attentional prioritization mechanism within WM aids flexible behavior in a dynamic world where there are multiple relevant items required for the task in hand whose relative priority changes frequently.

Recently, retro-cue benefits have been claimed to reflect an increase in the accessibility of the cued item in the absence of sustained selective attention[83]. According to this idea, the cued item is first attended and selected in memory. Then, its status is reconfigured in a way to make it more accessible for behavior. After this reconfiguration is complete, there is no need for sustained attention to keep the previously attended item in a prioritized accessible state. This theoretical model is in line with the pattern of results in 50% valid blocks of the present study where selective attention to the cued item was not sustained yet there were behavioral benefits for the cued item. It is thus possible that the cued item was reconfigured for accessibility without sustained attention. However, attention was sustained till the end of the trial in 80% valid blocks. Our results therefore show that, while a brief attentional selection might be sufficient for increasing the accessibility of a task-relevant item, highly relevant items are attended in a sustained manner.

Several studies have claimed that unattended items are stored silently, without sustained neural activity[68,76–79]. According to the activity-silent model of WM, unattended representations are stored through patterns of synaptic weights. Contrary to the activity-silent model of unattended WM items, we find in our experiment that unattended items are stored actively following retro-cues of low reliability, as reflected in the CDA. One important difference between the current study and those that support activity-silent WM is that here we used probabilistic retro-cues that did not provide certainty regarding which item was going to be tested. It is possible that when the retro-cue is 100% predictive of which item is going to be tested first, which was the case for the studies in support of activity-silent WM[68,79], the temporarily irrelevant item might be stored in an

bioRxiv preprint doi: https://doi.org/10.1101/320952; this version posted May 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

activity-silent state that relies on synaptic weights. Later, when this information is relevant again, it can then be retrieved from its passive state and brought back to an active state. On the other hand, if the retro-cue is not highly reliable, all items might be kept in an active state but with different attentional priorities. Consistent with this claim, here we show that an active WM trace for unattended items is present only when the retro-cue is less reliable. Another, but not mutually exclusive explanation for the absence of persistent activity for unattended items in studies that support activity-silent WM is that different cortical regions are responsible for storing attended and unattended items in WM. Recently Christophel et al. (2018) showed that visual cortex maintains only attended items while intraparietal areas and the frontal eye fields maintain both attended and unattended items[80]. This finding is consistent with our results given that the parietal cortex and frontal eye fields have been proposed as neural origins of the CDA[61,81,82], the ERP index we used in the current study to show the active maintenance of items in WM.

CDA has been traditionally defined as a sustained relative negativity contralateral to the memory items presented on one hemifield of the screen while the other hemifield is ignored. Here, the memory display contained memory items on both hemifield. Thus, we hypothesized that the emergence of a CDA following a retro-cue would mean that the item contralateral to the retro-cue is continued to be stored while item ipsilateral to the retro-cue is dropped[83]. An alternative explanation for the emergence of the CDA is that it reflects a boost for the cued item instead of signaling the loss of the non-cued item.[84] Given that the CDA reflects storage of the items in the contralateral side[64], an impact on the storage of the cued item should be reflected on the signal contralateral to the *cued item*. However, contrary to this alternative explanation, the retro-cue reliability affected the signal contralateral to the *non-cued item*. This result suggests that the CDA in the present study was a result of dropping the non-cued item instead of a mnemonic boost for the cued item.[57,85]

It has been argued that the CDA and contralateral alpha suppression are strongly related signals, to the extent that the CDA reflects the envelope of the alpha band power asymmetry between hemifields[86]. This would mean that the CDA and contralateral alpha suppression reflect one and the same attentional mechanism. However, this conclusion was based on an experiment that manipulated neither WM load nor the task demands, thus confounding task relevance and storage. In the present study, by having multiple items with different task-relevance, we show that contralateral alpha suppression and CDA behaved differently across time. Moreover, we did not observe any correlation between the CDA and contralateral alpha suppression across participants. Lastly, the retro-cue reliability effect was reflected in differences in the signal ipsilateral to the cued item for the CDA, but contralateral to the cued item for the contralateral alpha suppression. Together, these results strongly suggest that CDA is not simply a reflection of lateral alpha power asymmetries. We are not the first to observe dissociable patterns of lateral alpha power asymmetry and CDA. Previous studies that manipulated task demands observed larger contralateral alpha suppression when a WM item was stored for a more demanding task while the CDA was unchanged[52,87]. Here we extent these findings by showing both equal CDA

bioRxiv preprint doi: https://doi.org/10.1101/320952; this version posted May 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

and different contralateral alpha power, and also equal contralateral alpha power and different CDAs at different time points within the same dataset, thus provide a stronger evidence for a dissociation between these two signals. Together, these findings suggest that the CDA reflects storage in WM and the contralateral alpha suppression reflects allocation of attention within WM[88] and argue against a recent claim that suggested CDA reflects the current focus of attention instead of storage in WM[89].

In sum, by manipulating the reliability of retro-cues that indicate which of multiple WM items is most likely to be tested, we show that unattended items were kept in WM, but only when there was a relatively high chance that they could later be tested. Thus, we propose that the decision to drop an item from WM is separate than the decision to allocate attention away from it, and that these decisions can be flexibly adjusted based on dynamic changes in the relative importance of WM representations for the task in hand.

## Author Contributions

EG, DvM, MM, and CO designed the study. EG programmed the experiment. EG and KD collected the data. EG analyzed the data. EG, DvM, JF, MM, and CO wrote the manuscript.

## Acknowledgments

This work was supported by de Nederlandse organisatie voor Wetenschappelijk Onderzoek (The Netherlands Organisation for Scientific Research) to MM and C.O. (grantnumber 404-10-004). We would like to thank Joram van Driel and Ingmar de Vries for helpful discussions.

bioRxiv preprint doi: https://doi.org/10.1101/320952; this version posted May 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

# References


1. Carpenter, P. A., Just, M. A. & Shell, P. What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review* **97,** 404–431 (1990).
2. Fukuda, K., Vogel, E., Mayr, U. & Awh, E. Quantity, not quality: the relationship between fluid intelligence and working memory capacity. *Psychonomic Bulletin & Review* **17,** 673–679 (2010).
3. Kane, M. J. & Engle, R. W. The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective. *Psychonomic Bulletin & Review* **9,** 637–671 (2002).
4. Luck, S. J. & Vogel, E. K. Visual working memory capacity: from psychophysics and neurobiology to individual differences. *Trends in Cognitive Sciences* **17,** 391–400 (2013).
5. Bays, P. M. & Husain, M. Dynamic Shifts of Limited Working Memory Resources in Human Vision. *Science* **321,** 851 (2008).
6. Luck, S. J. & Vogel, E. K. The capacity of visual working memory for features and conjunctions. *Nature* **390,** 279 (1997).
7. Vogel, E. K., McCollough, A. W. & Machizawa, M. G. Neural measures reveal individual differences in controlling access to working memory. *Nature* **438,** 500 (2005).
8. Cusack, R., Lehmann, M., Veldsman, M. & Mitchell, D. J. Encoding strategy and not visual working memory capacity correlates with intelligence. *Psychonomic Bulletin & Review* **16,** 641–647 (2009).
9. LaBar, K. S., Gitelman, D. R., Parrish, T. B. & Mesulam, M.-M. Neuroanatomic Overlap of Working Memory and Spatial Attention Networks: A Functional MRI Comparison within Subjects. *NeuroImage* **10,** 695–704 (1999).
10. Engle, R. W., Kane, M. J. & Tuholski, S. W. Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and functions of the prefrontal cortex. *Models of working memory: Mechanisms of active maintenance and executive control.* 102–134 (1999). doi:10.1017/CBO9781139174909.007
11. Kane, M. J., Bleckley, M. K., Conway, A. R. A. & Engle, R. W. A controlled-attention view of working-memory capacity. *Journal of Experimental Psychology: General* **130,** 169–183 (2001).
12. Curtis, C. & D'Esposito, M. Persistent activity in the prefrontal cortex during working memory. *Trends in Cognitive Sciences* **7,** 415–423 (2003).
13. Corbetta, M., Kincade, J. M. & Shulman, G. L. Neural Systems for Visual Orienting and Their Relationships to Spatial Working Memory. *Journal of Cognitive Neuroscience* **14,** 508–523 (2002).
14. Kiyonaga, A. & Egner, T. Working memory as internal attention: Toward an integrative account of internal and external selection processes. *Psychonomic Bulletin & Review* **20,** 228–242 (2013).
15. Postle, B. R. Working memory as an emergent property of the mind and brain. *Neuroscience* **139,** 23–38 (2006).
16. Theeuwes, J., Belopolsky, A. & Olivers, C. N. L. Interactions between working memory, attention and eye movements. *Acta Psychologica* **132,** 106–114 (2009).
17. Awh, E. & Jonides, J. Overlapping mechanisms of attention and spatial working memory. *Trends in Cognitive Sciences* **5,** 119–126 (2001).
18. Awh, E. & Jonides, J. Spatial working memory and spatial selective attention. in *The attentive brain.* 353–380 (The MIT Press, 1998).
19. Gazzaley, A. & Nobre, A. C. Top-down modulation: bridging selective attention and working memory. *Trends in Cognitive Sciences* **16,** 129–135 (2012).
20. Awh, E., Anllo-Vento, L. & Hillyard, S. A. The Role of Spatial Selective Attention in Working Memory for Locations: Evidence from Event-Related Potentials. *Journal of Cognitive Neuroscience* **12,** 840–847 (2000).
21. Awh, E., Vogel, E. K. & Oh, S.-H. Interactions between attention and working memory. *Neuroscience* **139,** 201–208 (2006).
22. Olivers, C. N. L., Peters, J., Houtkamp, R. & Roelfsema, P. R. Different states in visual working memory: when it guides attention and when it does not. *Trends in Cognitive Sciences* **15,** 327–334 (2011).
23. Oberauer, K. Access to information in working memory: Exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **28,** 411–421 (2002).
24. Rerko, L. & Oberauer, K. Focused, unfocused, and defocused information in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **39,** 1075–1096 (2013).

bioRxiv preprint doi: https://doi.org/10.1101/320952; this version posted May 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

25. LaRocque, J. J., Lewis-Peacock, J. A. & Postle, B. R. Multiple neural states of representation in short-term memory? It's a matter of attention. *Frontiers in Human Neuroscience* **8,** 5 (2014).
26. Cowan, N. Attention and memory: An integrated framework. *Attention and memory: An integrated framework.* xv, 321–xv, 321 (1995).
27. Nee, D. E. & Jonides, J. Common and distinct neural correlates of perceptual and memorial selection. *NeuroImage* **45,** 963–975 (2009).
28. Nobre, A. C. *et al.* Orienting Attention to Locations in Perceptual Versus Mental Representations. *Journal of Cognitive Neuroscience* **16,** 363–373 (2004).
29. Tamber-Rosenau, B. J., Esterman, M., Chiu, Y.-C. & Yantis, S. Cortical Mechanisms of Cognitive Control for Shifting Attention in Vision and Working Memory. *Journal of Cognitive Neuroscience* **23,** 2905–2919 (2011).
30. Rerko, L. & Oberauer, K. Focused, unfocused, and defocused information in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **39,** 1075–1096 (2013).
31. Astle, D. E., Summerfield, J., Griffin, I. & Nobre, A. C. Orienting attention to locations in mental representations. *Attention, Perception, & Psychophysics* **74,** 146–162 (2012).
32. Myers, N. E., Chekroud, S. R., Stokes, M. G. & Nobre, A. C. Benefits of Flexible Prioritization in Working Memory Can Arise Without Costs. *Journal of Experimental Psychology: Human Perception and Performance* No Pagination Specified-No Pagination Specified (2017). doi:10.1037/xhp0000449
33. Griffin, I. C. & Nobre, A. C. Orienting Attention to Locations in Internal Representations. *Journal of Cognitive Neuroscience* **15,** 1176–1194 (2003).
34. Lepsien, J., Griffin, I. C., Devlin, J. T. & Nobre, A. C. Directing spatial attention in mental representations: Interactions between attentional orienting and working-memory load. *NeuroImage* **26,** 733–743 (2005).
35. Sligte, I. G., Scholte, H. S. & Lamme, V. A. F. Are There Multiple Visual Short-Term Memory Stores? *PLOS ONE* **3,** e1699 (2008).
36. Matsukura, M., Luck, S. J. & Vecera, S. P. Attention effects during visual short-term memory maintenance: Protection or prioritization? *Perception & Psychophysics* **69,** 1422–1434 (2007).
37. Pertzov, Y., Bays, P. M., Joseph, S. & Husain, M. Rapid forgetting prevented by retrospective attention cues. *Journal of Experimental Psychology: Human Perception and Performance* **39,** 1224–1231 (2013).
38. van Moorselaar, D., Olivers, C. N. L., Theeuwes, J., Lamme, V. A. F. & Sligte, I. G. Forgotten but not gone: Retro-cue costs and benefits in a double-cueing paradigm suggest multiple states in visual short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **41,** 1755–1763 (2015).
39. Landman, R., Spekreijse, H. & Lamme, V. A. F. Large capacity storage of integrated objects before change blindness. *Vision Research* **43,** 149–164 (2003).
40. Gunseli, E., van Moorselaar, D., Meeter, M. & Olivers, C. N. L. The reliability of retro-cues determines the fate of noncued visual working memory representations. *Psychonomic Bulletin & Review* **22,** 1334–1341 (2015).
41. Lepsien, J. & Nobre, A. C. Attentional Modulation of Object Representations in Working Memory. *Cerebral Cortex* **17,** 2072–2083 (2007).
42. Pertzov, Y., Manohar, S. & Husain, M. Rapid forgetting results from competition over time between items in visual working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **43,** 528–536 (2017).
43. Makovski, T., Sussman, R. & Jiang, Y. V. Orienting attention in visual working memory reduces interference from memory probes. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **34,** 369–380 (2008).
44. van Moorselaar, D., Gunseli, E., Theeuwes, J. & N. L. Olivers, C. The time course of protecting a visual memory representation from perceptual interference. *Frontiers in Human Neuroscience* **8,** 1053 (2015).
45. Souza, A. S., Rerko, L. & Oberauer, K. Getting more from visual working memory: Retro-cues enhance retrieval and protect from visual interference. *Journal of Experimental Psychology: Human Perception and Performance* **42,** 890–910 (2016).
46. Wang, B., Theeuwes, J. & Olivers, C. N. L. When Shorter Delays Lead to Worse Memories: Task Disruption Makes Visual Working Memory Temporarily Vulnerable to Test Interference. *Journal of Experimental Psychology: Learning, Memory, and Cognition* No Pagination Specified-No Pagination Specified (2017). doi:10.1037/xlm0000468
47. Zhang, W. & Luck, S. J. Discrete fixed-resolution representations in visual working memory. *Nature* **453,** 233 (2008).

bioRxiv preprint doi: https://doi.org/10.1101/320952; this version posted May 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

48. Klimesch, W. Alpha-band oscillations, attention, and controlled access to stored information. *Trends in Cognitive Sciences* **16,** 606–617 (2012).
49. Myers, N. E., Walther, L., Wallis, G., Stokes, M. G. & Nobre, A. C. Temporal Dynamics of Attention during Encoding versus Maintenance of Working Memory: Complementary Views from Event-related Potentials and Alpha-band Oscillations. *Journal of Cognitive Neuroscience* **27,** 492–508 (2014).
50. Poch, C., Campo, P. & Barnes, G. R. Modulation of alpha and gamma oscillations related to retrospectively orienting attention within working memory. *Eur J Neurosci* **40,** 2399–2405 (2014).
51. Poch, C., Capilla, A., Hinojosa, J. A. & Campo, P. Selection within working memory based on a color retro-cue modulates alpha oscillations. *Neuropsychologia* **106,** 133–137 (2017).
52. de Vries, I. E. J., van Driel, J. & Olivers, C. N. L. Posterior α EEG Dynamics Dissociate Current from Future Goals in Working Memory-Guided Visual Search. *J. Neurosci.* **37,** 1591 (2017).
53. Foster, J. J., Sutterer, D. W., Serences, J. T., Vogel, E. K. & Awh, E. The topography of alpha-band activity tracks the content of spatial working memory. *Journal of Neurophysiology* **115,** 168–177 (2015).
54. Ikkai, A., McCollough, A. W. & Vogel, E. K. Contralateral Delay Activity Provides a Neural Measure of the Number of Representations in Visual Working Memory. *J Neurophysiol* **103,** 1963 (2010).
55. Klaver, P., Talsma, D., Wijers, A. A., Heinze, H.-J. & Mulder, G. An event-related brain potential correlate of visual short-term memory. *NeuroReport* **10,** (1999).
56. Vogel, E. K. & Machizawa, M. G. Neural activity predicts individual differences in visual working memory capacity. *Nature* **428,** 748 (2004).
57. Williams, M. & Woodman, G. F. Directed Forgetting and Directed Remembering in Visual Working Memory. *Journal of experimental psychology. Learning, memory, and cognition* **38,** 10.1037/a0027389 (2012).
58. Bays, P. M., Catalao, R. F. G. & Husain, M. The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision* **9,** 7–7 (2009).
59. Suchow, J. W., Brady, T. F., Fougnie, D. & Alvarez, G. A. Modeling visual working memory with the MemToolbox. *Journal of Vision* **13,** 9–9 (2013).
60. Delorme, A. & Makeig, S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods* **134,** 9–21 (2004).
61. Luria, R., Balaban, H., Awh, E. & Vogel, E. K. The contralateral delay activity as a neural measure of visual working memory. *Neuroscience and biobehavioral reviews* **62,** 100–108 (2016).
62. Adam, K. C. S., Robison, M. K. & Vogel, E. K. Contralateral Delay Activity Tracks Fluctuations in Working Memory Performance. *Journal of Cognitive Neuroscience* 1–12 (2018). doi:10.1162/jocn_a_01233
63. Maris, E. & Oostenveld, R. Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods* **164,** 177–190 (2007).
64. Cohen, M. X. *Analyzing Neural Time Series Data*. (MIT Press, 2014).
65. Arend, A. M. & Zimmer, H. D. What Does Ipsilateral Delay Activity Reflect? Inferences from Slow Potentials in a Lateralized Visual Working Memory Task. *Journal of Cognitive Neuroscience* **23,** 4048–4056 (2011).
66. Postle, B. R. Working memory as an emergent property of the mind and brain. *Neuroscience* **139,** 23–38 (2006).
67. Awh, E. & Jonides, J. Spatial working memory and spatial selective attention. in *The attentive brain.* 353–380 (The MIT Press, 1998).
68. Wolff, M. J., Jochim, J., Akyürek, E. G. & Stokes, M. G. Dynamic hidden states underlying working-memory-guided behavior. *Nature Neuroscience* **20,** 864 (2017).
69. Sprague, T. C., Ester, E. F. & Serences, J. T. Restoring latent visual working memory representations in human cortex. *Neuron* **91,** 694–707 (2016).
70. Erickson, M. A., Maramara, L. A. & Lisman, J. A Single Brief Burst Induces GluR1-dependent Associative Short-term Potentiation: A Potential Mechanism for Short-term Memory. *Journal of Cognitive Neuroscience* **22,** 2530–2540 (2009).
71. Jafarpour, A., Penny, W., Barnes, G., Knight, R. T. & Duzel, E. Working Memory Replay Prioritizes Weakly Attended Events. *eNeuro* **4,** ENEURO.0171-17.2017 (2017).
72. Marian E. Berryhill, Lauren L. Richmond, Cara S. Shay & Ingrid R. Olson. Shifting Attention among Working Memory Representations: Testing Cue Type, Awareness, and Strategic Control. *Quarterly Journal of Experimental Psychology* **65,** 426–438 (2012).

bioRxiv preprint doi: https://doi.org/10.1101/320952; this version posted May 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.


73. Gözenman, F., Tanoue, R. T., Metoyer, T. & Berryhill, M. E. Invalid retro-cues can eliminate the retro-cue benefit: Evidence for a hybridized account. *Journal of experimental psychology. Human perception and performance* **40,** 1748–1754 (2014).
74. McElree, B. Working Memory and Focal Attention. *Journal of experimental psychology. Learning, memory, and cognition* **27,** 817–835 (2001).
75. Cowan, N. *Attention and memory : an integrated framework*. (Oxford University Press ; Clarendon Press, 1995).
76. Stokes, M. G. 'Activity-silent' working memory in prefrontal cortex: a dynamic coding framework. *Trends in Cognitive Sciences* **19,** 394–405 (2015).
77. Mongillo, G., Barak, O. & Tsodyks, M. Synaptic Theory of Working Memory. *Science* **319,** 1543 (2008).
78. Barak, O. & Tsodyks, M. Working models of working memory. *Current Opinion in Neurobiology* **25,** 20–24 (2014).
79. Rose, N. S. *et al.* Reactivation of latent working memories with transcranial magnetic stimulation. *Science* **354,** 1136 (2016).
80. Christophel, T. B., Iamshchinina, P., Yan, C., Allefeld, C. & Haynes, J.-D. Cortical specialization for attended versus unattended working memory. *Nature Neuroscience* (2018). doi:10.1038/s41593-018-0094-4
81. Robitaille, N., Grimault, S. & Jolicœur, P. Bilateral parietal and contralateral responses during maintenance of unilaterally encoded objects in visual short-term memory: Evidence from magnetoencephalography. *Psychophysiology* **46,** 1090–1099 (2009).
82. Becke, A., Müller, N., Vellage, A., Schoenfeld, M. A. & Hopf, J.-M. Neural sources of visual working memory maintenance in human parietal and ventral extrastriate visual cortex. *NeuroImage* **110,** 78–86 (2015).
83. Myers, N. E., Stokes, M. G. & Nobre, A. C. Prioritizing Information during Working Memory: Beyond Sustained Internal Attention. *Trends in Cognitive Sciences* **21,** 449–461
84. Heuer, A. & Schubö, A. The Focus of Attention in Visual Working Memory: Protection of Focused Representations and Its Individual Variation. *PLoS ONE* **11,** e0154228 (2016).
85. Kuo, B.-C., Stokes, M. G. & Nobre, A. C. Attention Modulates Maintenance of Representations in Visual Short-Term Memory. *Journal of cognitive neuroscience* **24,** 51–60 (2012).
86. van Dijk, H., van der Werf, J., Mazaheri, A., Medendorp, W. P. & Jensen, O. Modulations in oscillatory activity with amplitude asymmetry can produce cognitively relevant event-related responses. *Proc Natl Acad Sci USA* **107,** 900 (2010).
87. van Driel, J., Gunseli, E., Meeter, M. & Olivers, C. N. L. Local and interregional alpha EEG dynamics dissociate between memory for search and memory for recognition. *NeuroImage* **149,** 114–128 (2017).
88. Bae, G.-Y. & Luck, S. J. Dissociable Decoding of Spatial Attention and Working Memory from EEG Oscillations and Sustained Potentials. *J. Neurosci.* **38,** 409 (2018).
89. Berggren, N. & Eimer, M. Does Contralateral Delay Activity Reflect Working Memory Storage or the Current Focus of Spatial Attention within Visual Working Memory? *Journal of Cognitive Neuroscience* **28,** 2003–2020 (2016).