Attention, Perception, & Psychophysics
https://doi.org/10.3758/s13414-023-02724-2

Are irrelevant items actively deleted from visual working memory? No evidence from repulsion and attraction effects in dual-retrocue tasks

Joshua P. Rhilinger[1] · **Chenlingxi Xu**[1] · **Nathan S. Rose**[1]

Accepted: 26 April 2023

Abstract

Some theories propose that working memory (WM) involves the active deletion of irrelevant information, including items that were retained in WM, but are no longer relevant for ongoing cognition. Considerable evidence suggests that active-deletion occurs for categorical representations, but whether it also occurs for recall of features that are typically bound together in an object, such as line orientations, is unclear. In two experiments, with or without binding instructions, healthy young adults maintained two orientations, focused attention to recall the orientation cued first, and then switched attention to recall the orientation cued second, at which point the uncued orientation was no longer relevant on the trial. In contrast to the active-deletion hypothesis, the results showed that the no-longer-relevant items exerted the strongest bias on participants' recall, which was either repulsive or attractive depending on the degree of difference between the target and nontarget orientations and the proximity to cardinal axes. We suggest that visual WM binds features like line orientations into chunked representations, and an irrelevant feature of a chunked object cannot be actively deleted – it biases recall of the target feature. Models of WM need to be updated to explain this dynamic phenomenon.

Keywords Working memory · Deletion · Removal · Suppression · Bias · Repulsion · Attraction

Introduction

Working memory (WM) is used to maintain and manipulate items in mind while using them to accomplish a goal (Baddeley, 2012). Many theorists propose that, in addition to actively retaining goal-relevant information, WM also involves actively deleting information that is no longer relevant[1] (Hasher et al., 2007; Lewis-Peacock et al., 2018; Oberauer, 2009). However, the results reported here suggest that this mechanism may not be universally applied.

Measuring the prioritization and deletion of items in working memory (WM) with retrocue tasks

While maintaining goal-relevant information in WM, it helps to be cued to the information that is most relevant for ongoing cognition (e.g., which item will be tested by an upcoming recall or recognition test) even if the cue comes after the information has been presented and encoded in WM (i.e., retrocues) (Souza & Oberauer, 2016). Numerous studies have shown differences in behavior (accuracy, response times) and brain activity (EEG/ERP, MEG, fMRI) associated with WM for retrocued versus uncued items (for a meta-analysis, see Wallis et al., 2015). Many theorists have concluded that these differences arise when internal attention selects and protects retrocued items against interference from other items in WM (Souza et al., 2016). Other theorists have posited that retrocueing benefits relevant items because controlled attentional processes can be strategically used to actively delete irrelevant items from WM (e.g., Lewis-Peacock et al., 2018). However, it is unclear whether cueing is beneficial because people selectively attend to and enhance the representation of relevant items, because they selectively

---

[1] There are many similar terms that have been used to describe this process including suppression, inhibition, deletion, removal, clearing, gating, interference resolution, etc. The extent to which these terms connote similar or different processes is unclear. Here we use the term "active deletion" to refer to the general mechanism, and discuss how clarification among related and distinct concepts is needed.

✉ Nathan S. Rose
nrose1@nd.edu

[1] University of Notre Dame, 390 Corbett Family Hall, Notre Dame, IN 46556, USA

delete irrelevant items, or because of both (Lintz & Johnson, 2021). Tasks with multiple retrocues are particularly revealing because the consequences of prioritizing one item over the rest can be assessed for both the initially cued and the initially uncued items.

Consider a situation in which two to-be-remembered items are presented and actively retained in WM; an initial retrocue indicates which item is to be tested first, and then a second retrocue indicates which item is to be tested second. Using such a task with categorical stimuli (faces, words, or directions of motion) as memoranda, Rose et al. (2016) showed behavioral and neural evidence that supported the idea that no-longer relevant items were actively deleted from WM. When the two items were initially presented and retained in WM, the category of both items could be decoded from the participant's brain activity using fMRI or EEG. Following the first retrocue, neural representation of the uncued item dropped to baseline as if it were no longer actively retained "in WM" – but it could be reactivated by a single pulse of transcranial magnetic stimulation (TMS) applied to a category-selective region of posterior cortex. This suggested that, while this uncued item was still potentially relevant later on in the trial, the uncued item was passively retained in WM via "activity-silent" (short-term synaptic plasticity) mechanisms (Rose, 2020; Silvanto, 2017).[2] However, following the second retrocue, which indicated that the uncued item was no longer relevant on the trial, TMS could no longer reactivate the uncued (no-longer-relevant) item. This suggested that the no-longer-relevant item was actively deleted from WM following the second retrocue. For replications and extensions, see Fulvio and Postle (2020) and Wolff et al. (2017).

Related research has also shown evidence for an active-deletion process that removes items cued as no longer relevant for WM (Oberauer, 2018; for a review, see Lewis-Peacock et al., 2018). However, other evidence suggests that this active-deletion mechanism is not always utilized (Dagry et al., 2017; Dagry & Barrouillet, 2017; Lilienthal et al., 2015; Lintz & Johnson, 2021; Oberauer, 2018). Although the active-deletion hypothesis proposes that slower presentation rates allow more time to remove distractors and also that deleted items should be less accessible on subsequent memory tests, contradictory evidence has been shown from repetition priming, lexical decision, and subsequent memory effects of distractors (e.g., Dagry et al., 2017; Dagry & Barrouillet, 2017; Lilienthal et al., 2015), even when participants are explicitly instructed to either remove uncued items or refresh cued items (Lintz & Johnson, 2021).

Revealing the activation state and interference among items in WM with repulsion and attraction effects

Another way to test how items are retained in (or deleted from) WM is to examine the extent to which retained items interfere with one another during recall (Wildegger et al., 2015). For stimuli that are represented in a continuous feature space such as orientations, spatial locations, colors, etc., it is possible to detect subtle numerical biases between items retained in WM (Bae & Luck, 2017). For example, when attempting to recall a cued item in WM (e.g., an orientation of 30°), an uncued item in WM (e.g., an orientation of 10°) can systematically bias recall of the cued item either away from the uncued item – a phenomenon called "repulsion" (Kiyonaga & Egner, 2016) – or toward the uncued item – a phenomenon called "attraction" (Chunharas et al., 2022). In this example, repulsion would be reflected by the participant recalling the target orientation to be farther in the feature space from the distractor (e.g., 32° vs. 28°). Such biases have also been shown to influence WM for color (Golomb, 2015), motion (Czoschke et al., 2019), and even faces whose features vary along continua (Mallett et al., 2020). Bias can come from both task-irrelevant distractors or memoranda from previous trials, as in the so-called serial dependence effect (Shan & Postle, 2022). The phenomenon is consistent with neurocomputational models of visual WM that posit repulsive bias between similar items due to lateral inhibition (Johnson et al., 2009), which can flip in sign from an attractive bias from a previous trial to a repulsive bias within a trial (Fritsche et al., 2020).

The present study

The present study used orientations and a similar double-retrocue paradigm to that used by Rose et al. (2016) to examine repulsion and attraction effects of competing memory items (Fig. 1).

In a double-retrocue task, bias should be largest on recall 1 (when the uncued item is still potentially relevant on the trial). If the uncued item is actively deleted after the second cue (because it is no longer relevant on the trial), then bias should be smaller on recall 2 than on recall 1. However, larger bias from the uncued item on recall 2 compared to recall 1 would suggest that the no-longer-relevant item persists in WM; such findings would question the generalizability of the active-deletion mechanism.

Our task design and analyses allowed the measurement of memory fidelity of items held in such cued or uncued states, as well as the relative contributions of distinct sources of influence on their recall. The use of continuous stimuli enabled us to explore the effects of dropping an uncued memory

---

2FL01 [2] For alternative interpretations, see Schneegans and Bays (2017) and
2FL02 Stokes et al. (2020).

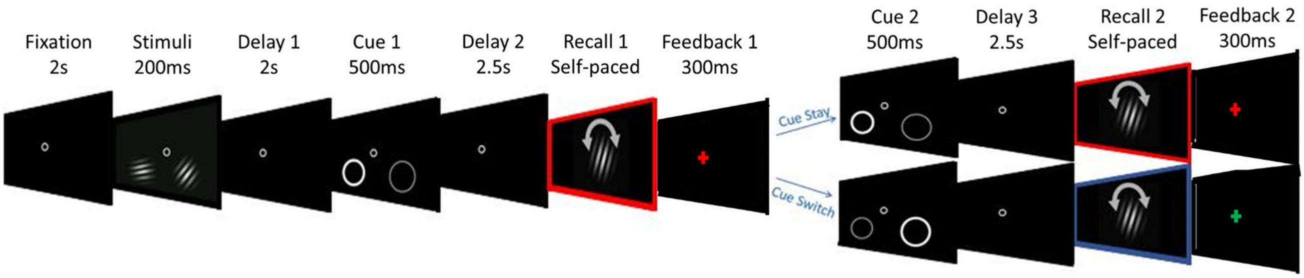

Fig. 1 Double-retrocue task design. Gabor-stimuli were simultaneously presented in the lower-right and lower-left visual hemifield. Following a delay, a bold white outline at the stimulus location served as the retrocue (with 100% validity). Following delay 2, a random Gabor patch was presented at central fixation and participants rotated the patch to match the cued orientation. After submitting their response and receiving feedback, either the same stimulus was cued again ("stay" trial), or the originally uncued-stimulus was cued ("switch" trial) for the second recall test and feedback. Note that feedback was provided by turning the fixation cross green, yellow, or red for 15°, 15–30°, or >30° of error, respectively. The red and blue borders were not shown; they depict the recall-1, recall-2-stay, and recall-2-switch conditions, respectively

item from an attended state on memory, which is more sensitive to detecting subtle biases and the sources of variability in recall than other paradigms such as recognition of categorical stimuli. Specifically, we used computational modeling to separate memory errors into precision (defined as the standard deviation of errors), guess rate (defined as the likelihood that the participant had no memory trace for the target item), and swap error rate (defined as the likelihood a response reflects the uncued memory item; also known as a binding error) (Peters et al., 2019). We predicted that precision would be worse and guess and swap error rates would be higher for recall-2-switch trials, in which the initially uncued item was cued for recall on the second test, compared to recall-1 and recall-2-stay trials. Bias analyses were examined with the mixture model results in an attempt to elucidate the source of differences between cued and uncued items on these parameters. We had no a priori hypotheses about the direction of bias (repulsion or attraction) or differences between the conditions. The main hypothesis regarding bias was that if no-longer-relevant items were actively deleted, then bias should be less on recall 2 than on recall 1 responses. Any contrary evidence would call for a revision to the active-deletion hypothesis.

Experiment 1

Method

Participants Forty-one (M_{age} = 19.1 years, range = 18–35 years, 27 female) right-handed students with normal or corrected-to-normal vision were recruited to participate in the experiment. Participants provided informed consent (Institutional Review Board (IRB) protocol 17-02-3629) and were remunerated with cash or course credit (US$15 or 1 credit/h). Data for six participants were unavailable (three withdrew after screening, three due to technical errors); analyses were conducted on the remaining 35 participants' data.[3]

WM task Participants were seated approximately 37 cm (n = 19) or 57 cm (n = 16) away from a 24-in. ASUS computer monitor with 1,920 × 1,080 resolution and a 60-Hz refresh rate. The task and stimuli were generated and run in MATLAB using the Psychophysics Toolbox V3.0 (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997). Responses were given using the "1" and "2" buttons on the T9 number pad of a standard QWERTY keyboard to freely rotate the presented recall stimulus counterclockwise or clockwise, respectively.

Stimulus details Central fixation was identified by a white circle with an outer radius of a quarter pixel and an inner radius of an eighth pixel. The experimental stimuli consisted of two sine-wave gratings (i.e., Gabor patches) with a diameter of 2°, spatial frequency of 2 cycles/°, a phase of 0, and a Michelson contrast of 100%. The orientations were separated into seven distinct orientation bins with centers of 13°, 39°, 65°, 91°, 117°, 143°, and 169°, with each bin containing the same number of orientations. For a given trial, orientations were selected pseudo-randomly from these bins with a jitter of ± 5°, and the two stimuli in a given trial varied by more than 10°.

Location of stimulus presentation in the lower left and right visual hemifields was matched to phosphene localizations acquired from participants in an ongoing TMS study to target the early visual cortex (V1/V2). The retrocues consisted of circular outlines surrounding the locations where the stimuli were presented. The cued item was outlined by a bold (.5°) white circle; the non-cued item was outlined by a

---

[3] Data from six of the participants who were assigned to the sham rTMS condition are included because their performance was unaffected by TMS.

non-bold (0.15)° light-grey circle. Presenting circles at the locations of both the cued and uncued items was necessary to avoid selectively "pinging" the cued item with a visual impulse (Wolff et al., 2017).

Phosphene localization procedure The locations where stimuli were presented was determined by an ongoing rTMS study in which a different set of participants first underwent a phosphene localization and thresholding procedure in a dark room to determine if they could reliably see a circular-shaped phosphene in the lower-right visual field when holding central-fixation from single-pulses of TMS applied to left, early-visual cortex (V1/V2). If so, the TMS intensity at which a phosphene was induced in five out of ten trials was determined following established procedures (Abrahamyan et al., 2011; Rademaker et al., 2017). Then TMS intensity was set to 110% of the phosphene threshold, single pulses were applied at the localized area, and, following each pulse, participants were instructed to use the computer mouse to trace an outline of the perceived phosphene onto the black computer screen with a gray, central-fixation cross using custom MATLAB/PsychToolbox code.

Following the drawing of at least ten outlines, each outline was fit to an ellipse using the *fitellipse* function, the centroid of each ellipse was calculated, and the median centroid value (in X and Y screen-pixel coordinates) was recorded. These coordinates were used to determine the location at which the center of the right Gabor-orientation-patch was presented for the WM task. The left Gabor patch was presented in the contralateral visual field from these coordinates; the right Gabor patch was presented in the mirroring side of the visual field. Therefore, stimuli locations were individually determined and unique for each participant in the ongoing rTMS study. For the purposes of this behavioral-only, control experiment, different participants were randomly matched to the stimuli-locations determined for participants who completed the phosphene localization task and the rTMS version of the experiment in order to assess the potential impact of stimulus-location variability on performance. That is, participants in this behavioral-only, control study did not receive TMS; the locations at which stimuli were presented for a participant were matched to those that were generated for a corresponding participant in the rTMS experiment.[4]

Task procedure A white fixation circle was presented on a black screen at the beginning of each trial for 2 s and remained on the screen throughout stimulus and cue presentation (Fig. 1). Gabor patches were presented for .2 s in the lower visual-hemifield, one in the right-hemifield and the other symmetrically mirrored in the left-hemifield according to the locations determined by the phosphene-localization procedure described in the preceding section. After a 2-s delay, the first retrocue was presented for .5 s. A 2.5-s delay followed the cue before a random Gabor orientation was presented in the center of the screen. Participants were instructed to rotate the orientation to match the orientation of the cued-stimulus. Once the response was submitted, feedback was displayed at central-fixation for 0.3 s. Responses that were within 15°, 15–30°, or >30° away from the target turned the fixation cross green, yellow, or red, respectively.[5] Following the first feedback, a second retrocue was displayed for .5 s. This retrocue could signal that either the same stimulus would be tested a second time (a "stay" trial) or that the originally uncued stimulus would be tested (a "switch" trial). Trials were balanced so that there was an equal number of stay and switch trials in each block. A random Gabor patch was once again presented in the center of the screen after the 2.5-s delay, and participants rotated the Gabor patch to match the stimulus cued by the second retrocue. Feedback was once again given following the second recall response. Each block consisted of 56 trials, and participants completed 2-3 blocks in each session.

Data quality checks For the average accuracy analysis, in order to identify potential outliers in the data, for each recall condition per participant, the errors were first converted to z-scores, and any z-score > 3 or < -3 was removed from the data set. Since these responses were significant outliers, they likely reflect cases in which participants had no memory representation for the target item and resorted to guessing. Therefore, removing these responses before the average accuracy analysis enabled us to get a more accurate measure of memory performance. A total of 1.3% of the responses was removed (212 out of a total of 15,770 responses), and no more than 14 trials were removed from any recall condition for an individual participant. The analysis of errors was conducted on the non-z-score converted data as the circular deviation of the recalled orientation from the target orientation in degrees. The *boxplot* function in R Studio was then used across all participants to determine any outliers in the dataset (defined

---

[4] The degree of visual angle varied from a mean of 10.6° from central fixation (Median = 10°, SD = 4.1°). To assess the impact of this variability on the data, we conducted several Pearson's correlations between the degree of visual angle and our dependent variables. There were no significant correlations between degree of visual angle and any of the dependent variables reported here (rs $< .16$, n.s.). There were significant negative correlations between visual angle and mean response times for recall 1 ($r = -.44$) and recall 2 ($r = -.38$), but response time analyses were not the focus of the hypotheses tested here, so variability in degree of visual angle was not considered further.

[5] While feedback effects on recall-2-stay trials could complicate the interpretation of the results, analyses comparing recall-1 and recall-2-stay trials showed that the effects of feedback were minimal and did not substantially alter interpretation of the observed biases (see Online Supplemental Material (OSM) Figs. 2 and 3).

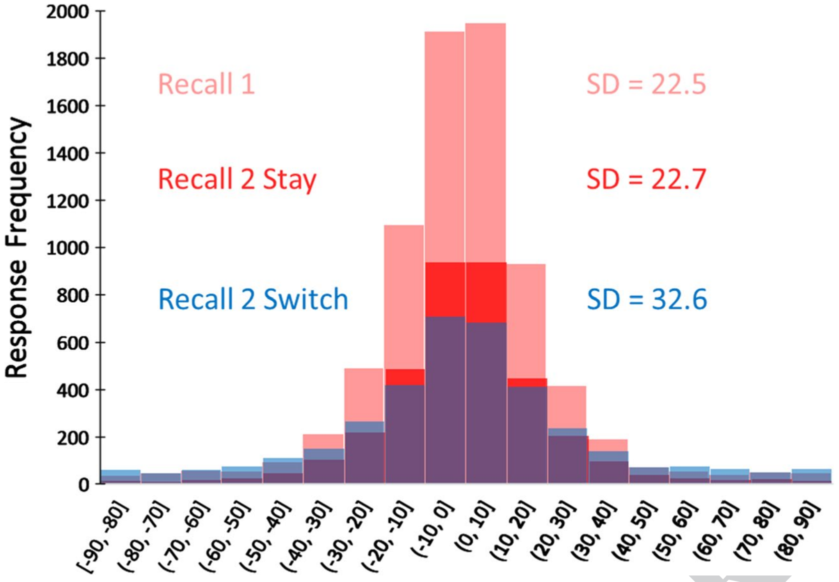

Fig. 2 Experiment 1: The frequency of recall errors and the standard deviation (SD, i.e., memory precision) in degrees relative to the target orientation for each condition (recall-1, recall-2-stay, and recall-2-switch) for all trials and all participants. Memory precision was similar for recall-1 and recall-2-stay trials, while recall-2-switch trials were less precise. Note that, by design, recall-1 had twice as many trials as recall-2-stay and recall-2-switch trials; also note the lack of any systematic bias to the left (negative degrees, i.e., counterclockwise) or right (positive degrees, i.e., clockwise) of the target orientation

as 1.5 times the interquartile range above or below the third or first quartiles, respectively); two participants were determined to be outliers and removed from the error analysis comparing behavioral performance across recall conditions. Thus, data from 33 participants were used in the behavioral analyses.

For the mixture model analyses, all trials for the remaining participants were included (even the trials previously identified by the z-score analysis as outliers) because the models attempted to separate errors by different parameters, so were able to account for outliers. One participant had an implausible recall-2-switch precision parameter (3.27E+28), suggesting that the mixture model failed to fit the data. Therefore, parameter values for this participant were not included in the group level analysis, leaving data from 32 participants to be included in the mixture model analyses. A mixed-design ANOVA showed that the interaction between performance on the three recall conditions and viewing distance was not significant ($F(2,62) = 0.71$, $p = 0.50$). Moreover, the correlations between performance and degrees of visual angle were not significant for any of the three recall conditions, *r*s = 0.05, 0.04, and 0.05, respectively, *p*s > 0.25. For the nontarget bias analysis, we included all trials, but removed the data of the two participants previously deemed outliers as in the average accuracy analysis.

Data analysis R Studio was used to perform all of the statistical tests on the model parameters and comparisons. The normality of the distribution of recall errors for each condition was assessed using a log10 transformation and Shapiro-Wilks tests, which confirmed normality (*p*s > 0.1442, see Fig. 2). For all analyses, two-tailed, Bonferroni-corrected t-tests were used, unless stated otherwise.

Errors were calculated as the circular difference between the target orientation and the response orientation, according to the von Mises distribution. The difference between the nontarget orientation (i.e., the uncued orientation) and the response was also calculated for the mixture modeling to determine the influence that the nontarget had on the response.

Mixture model analyses Mixture modeling was conducted on the errors using the MemToolbox (Suchow et al., 2013). The models were used to parameterize memory precision and the proportion of responses in which the participant likely guessed or committed a binding error. We plotted the response errors centered around the target response of 0 error. The Standard Mixture Model (Zhang & Luck, 2008) was compared to the Swap Model (Bays et al., 2009). The Standard Mixture Model used the distance of a response from the target value to determine both the probability that the error reflects the precision (reflected by *SD*) of the participant's memory for the target item and the probability that the response was a random guess (reflected by the uniform distribution called the guess rate, or *g* parameter). This model uses the following equation when fitting the data:

$$p(\hat{\theta}) = (1-\gamma)\phi_{\sigma}(\hat{\theta}-\theta) + \gamma\frac{1}{2\pi} \tag{1}$$

where θ serves as the target value (in radians), $\hat{\theta}$ serves as the response value, γ serves as the frequency of random guesses, and ϕ_σ serves as the circular analogue of the Von Mises distribution (*mean* = 0, *SD* = σ).

The Swap Model (Bays et al., 2009) includes the same precision and guess rate parameters as well as a third parameter, the swap error rate, which reflects the probability that a response reflects a memory for the nontarget item. By taking the accuracy of the response relative to the nontarget item into account, the swap error rate indicates the probability that the participant recalled the uncued item rather than the cued item. The Swap Model is described by the equation:

$$p(\hat{\theta}) = (1 - \gamma - \beta)\phi_\sigma(\hat{\theta} - \theta) + \gamma \frac{1}{2\pi} + \beta \frac{1}{m} \sum_{i}^{m} \phi_\sigma(\hat{\theta} - \theta_i^*) \tag{2}$$

where β serves as the probability of a swap error and $\{\theta_1{*}, \theta_2{*}, \ldots \theta_m{*}\}$ are the *m* nontarget line orientation values (Bays et al., 2009).

The responses for each recall condition (recall-1, recall-2-stay, and recall-2-switch) were modeled separately for each participant to see how memory changed when items were switched from an unprioritized- to a prioritized-state within-subjects design. The fits of each model were compared using the Akaike Information Criterion (AIC).

Nontarget bias analyses To assess the influence of the nontarget (uncued) orientation on recall of the target orientation we calculated and compared the amount of response bias relative to the nontarget on each of the three recall conditions for both the average amount of bias and as a function of the degree of difference between the target and nontarget orientations. Positive or negative error value indicates whether the response was biased away from the nontarget ("repulsion") or toward the nontarget ("attraction"), respectively. We performed Bonferroni-corrected t-tests on the average error relative to the nontarget across all trials to compare the amount of bias between the three recall conditions (using paired-sample t-tests) and for the average of bins of trials with small, medium, or large differences between the target and nontarget orientations (using one-sample t-tests vs. zero, i.e., no bias).

Results

The distributions of recall errors relative to the to-be-remembered target orientation on recall-1, recall-2-stay, and recall-2-switch trials for all participants are shown in Fig. 2.

To determine the consequences of holding information in an unprioritized state, we compared the average accuracies across the three recall conditions (recall-1, recall-2-stay, and recall-2-switch trials) as the absolute value of recall error (in degrees). Recall error was higher on recall-2-switch trials (*M* = 22.7, *SD* = 8.06) than on both recall-1 (*M* = 14.2, *SD* = 5.19; *t*(32) = -14.68, *p* < .001) and recall-2-stay trials (*M* = 14.6, *SD* = 5.77; *t*(32) = -13.50, *p* < .001), but there was no difference between recall-1 and recall-2-stay trials (*t*(32) = -0.69, *p* = 0.50) (OSM Fig. 1). These results support our hypotheses that shifting a memory item into an unattended state weakened the fidelity of memory for that item compared to items maintained in an attended state. We then conducted mixture model analyses on the data in order to better understand the source(s) of the differences.

Mixture modeling

Model preference We first compared the two mixture models to determine which of the models was a better fit to the data. We performed a Wilcoxon signed-rank test on the difference in the AIC values between the Standard Mixture Model and the Swap Model for all participants (as in Bays & Taylor, 2018). The Swap Model was preferred over the Standard Mixture Model for all three recall conditions: recall-1 (ΔM = 15.33, *p* < .001); recall-2-stay (ΔM = 6.77, *p* < .01); and recall-2-switch (ΔM = 9.45, *p* < .01).

Parameter differences To determine the effects of shifting attention between items, we compared the error parameters from the Swap Model across the three recall conditions (recall-1, recall-2-stay, and recall-2-switch). Three Bonferroni-corrected, two-tailed (unless stated otherwise) paired t-tests were performed for each parameter to compare all three recall conditions.

Precision As predicted, there was a statistically significant increase in the precision parameter (indicating less precision) on recall-2-switch trials compared with both recall-1 trials (*t*(31) = -5.64, *p* < .01) and recall-2-stay trials (*t*(31) = -5.61, *p* < .01). There was no significant difference in the precision parameter between recall-1 trials and recall-2-stay trials (*t*(31) = -0.70, *p* = 0.49, Fig. 3A).

Guess rate The guess rate was higher for recall-2-switch trials than for both recall-1 trials (*t*(31) = -7.34, *p* < .001) and recall-2-stay trials (*t*(31) = -5.26, *p* < .001), and there was no significant difference between recall-1 and recall-2-stay trials (*t*(31) = -1.94, *p* = 0.06, Fig. 3B).

Swap error rate As predicted, the estimated swap error rate was higher on recall-2-switch trials than on both recall-1 and recall-2-stay trials (*ts*(31) = -6.25 and -4.83, *ps* < .01 and .017, respectively);[6] the difference in swap error rate

---

[6] Note that these are one-tailed t-tests because of our a priori hypothesis that swap errors would increase on recall-2-switch trials.

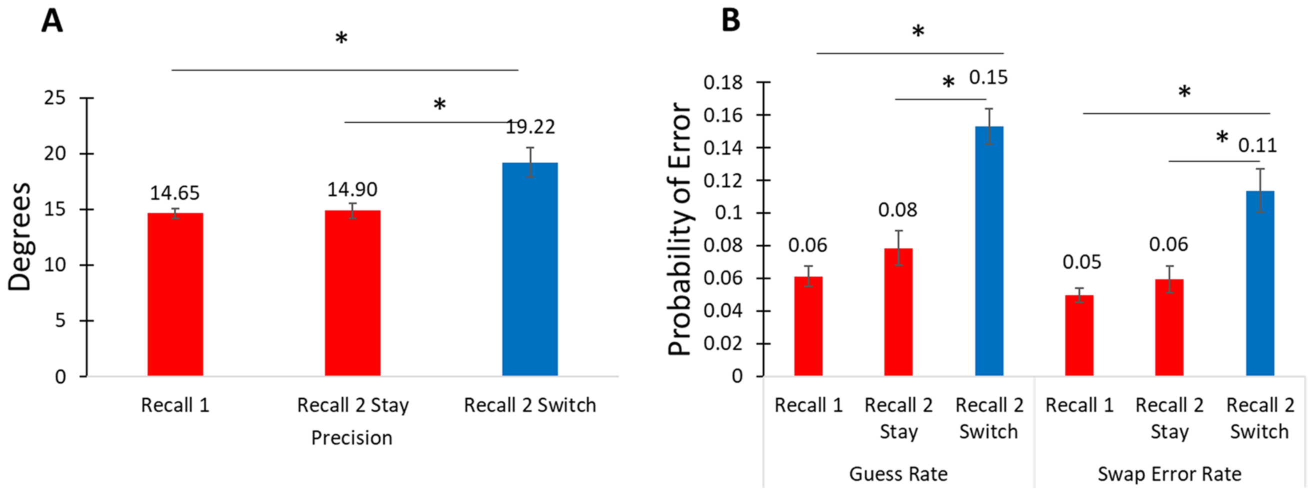

Fig. 3 Experiment 1: Average swap model parameter values. **A** The average precision parameter in degrees for the three recall conditions (lower values indicate better precision). Memory was less precise for recall-2-switch trials compared to both recall-1 and recall-2-stay trials. **B** The average guess rate and average swap error rate parameters for the three recall conditions (higher values indicate more guess or swap errors). Both the average guess rate and average swap error rates were higher for recall-2-switch trials compared to both recall-1 and recall-2-stay trials. Error bars reflect ±1 standard error of the mean; *p < 0.001

between recall-1 and recall-2-stay trials did not survive Bonferroni correction ($t(31) = -2.18$, $p = 0.04$, Fig. 3B).

Overall, these data support our hypotheses that holding an item in a deprioritized state results in worse memory fidelity for that item and also increases the commission of swap errors. This also increased the number of guesses.

Nontarget bias analysis: repulsion and attraction effects To elucidate the source of the differences between the recall conditions we investigated the role that the nontarget played in biasing response errors and how this bias changed across the recall conditions. We performed Bonferroni-corrected paired t-tests on the average error with bias across the three recall conditions. In contrast to the active-deletion hypothesis, there was no difference in repulsive bias between recall-1 and recall-2-stay trials ($t(32) = 0.15$, $p = 0.88$), and there was *less* repulsive bias on recall-1 than recall-2-switch trials ($t(32) = 3.52$, $p < 0.005$). There was also *more* repulsive bias on recall-2-switch versus stay trials despite the fact that the uncued item was no longer relevant during recall 2 for both stay and switch trials ($t(32) = -3.00$, $p < 0.01$, Fig. 4).

To further elucidate the source of bias on recall, we calculated the amount of bias as a function of the difference between the target and nontarget orientations. The purpose of this analysis was to assess whether the amount of bias that the uncued (nontarget) item had on recall of the cued (target) item depended on the similarity between the target and nontarget. Trials were binned around three orientation differences centered around relatively small (~25°), medium (~50°), and large (~75°) differences between the target and nontarget orientations.[7] The amount of bias on recall-1 and recall-2-stay trials was not significantly different from zero for the 25°, 50°, or 75° bins (*ps* > .05). For recall-2-switch trial, there was significant repulsion from the nontarget orientation when there were small or medium differences between the target and nontarget (*ps* < .01), but the amount of bias was not significant when there were large (~75°) differences between the target and the non-target ($p = .69$, Fig. 5). As discussed below, current neurocomputational models of visual WM posit that lateral inhibition mechanisms could drive repulsive bias seen between similar items (Johnson et al., 2009).

Are differences in precision, guess, and swap parameters due to bias? Finally, an exploratory correlational analysis was done to see if the poorer precision, guess, and swap error parameters on recall-2-switch trials that were observed (Fig. 3) were associated with the increase in repulsive bias that was observed on these trials (Figs. 4 and 5). Participants' precision parameter and their average bias on recall-2-switch trials were positively correlated ($r = .41$, $p = .02$), indicating that those with poorer memory precision had greater repulsive bias; in contrast, participants' guess and swap parameters were not correlated with their average

---

[7] Recall that the minimum difference between the target and non-target orientations was necessarily greater than 10. These three bin windows were created so that each bin represented the same range of orientation differences (e.g., the 25° bin includes trials in which the difference between the two orientations was between 14° and 36°), while ensuring an equal distribution of trial counts around the bin centers, and similar trial counts between the bins.

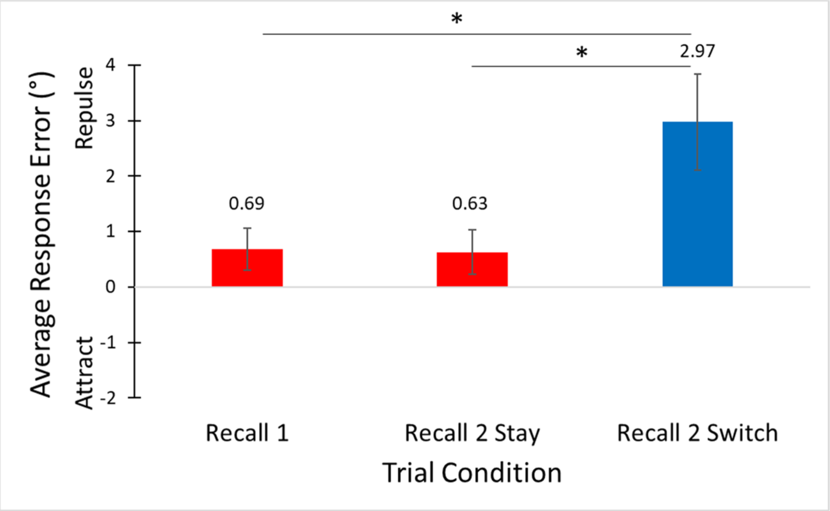

Fig. 4 Experiment 1: Average response error bias from the nontarget item for each trial condition. Responses were calculated based on whether errors were committed away from (greater than 0, i.e., repulsion) or closer to (less than 0, i.e., attraction) the orientation of the nontarget item and averaged for each trial condition. Average bias from the nontarget was not significantly different from 0 for recall 1 and recall-2-stay trials (*p*s>.05). Average bias was greater for recall-2-switch trials than for both recall-1 and recall-2-stay trials. Error bars reflect 1 SEM, $^*p < 0.001$

amount of bias (*r*s = .04 and -.29, *p*s = .83 and .11), indicating that the increase in guess and swap errors was not associated with the increase in bias on recall-2-switch trials. Also note that recall of nontargets (swap errors) would result in an attractive bias – not the observed repulsive bias that differed depending the degree of target-nontarget similarity.

Discussion

Compared to actively maintaining and recalling a cued item in WM, passively retaining and then returning an uncued item back into focal attention resulted in decreases in recall precision (which was associated with the degree of bias from

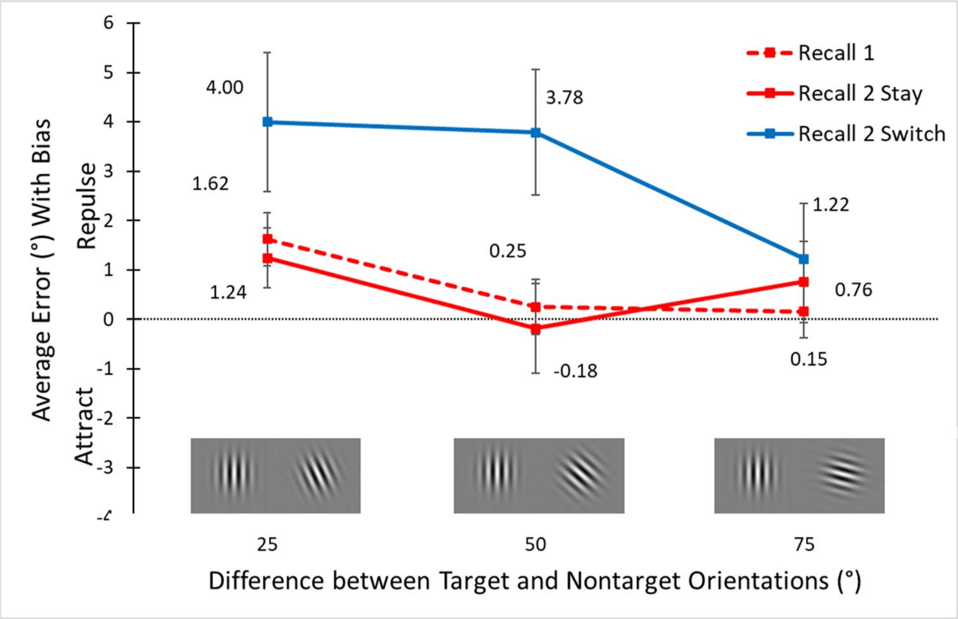

Fig. 5 Experiment 1: Average response error bias as a function of distance between orientation stimuli. Response errors were calculated as the degree of difference from the target orientation towards (negative) or away from (positive) the nontarget item, and the average response bias across participants was plotted for each recall condition. Bins were created using trials in which the difference between the stimuli were ± 11.5° from 25°, 50°, and 75°, respectively. In contrast to the active-deletion hypothesis, there was no difference in bias between recall-1 and recall-2-stay trials; recall-2-switch trials had significantly more bias than both recall-1 and recall-2-stay trials, especially when the target and nontarget orientations were more similar (see *Results* section). Error bars reflect ±1 standard error of the mean

the nontarget orientation), and increases in the probability that the participant guessed or recalled the nontarget item. These findings are consistent with hypotheses that internal attention can select one of multiple items in WM to prioritize its retention and recall over other items, and that items dropped from focal attention can be passively retained and reactivated when needed, via error-prone retrieval processes (see also LaRocque et al., 2015; Peters et al., 2019).

The key finding is that, contrary to the hypothesis that no-longer-relevant items are actively deleted from WM, these items persisted and biased recall of the target item held in focal attention, especially when the target and no-longer-relevant items were similar to one another. Moreover, recall-1 trials showed the same amount of bias as recall-2-stay trials and *less* bias than recall-2-switch trials, which contradicts the pattern predicted by the active-deletion hypothesis. Following the second retrocue, the uncued item was no longer relevant and, therefore, according to the active-deletion hypothesis, it should have been deleted from WM and should have resulted in less bias for responses on recall-2-stay and recall-2-switch trials. The present results suggest that no-longer-relevant items were not deleted from WM following the second retrocue.

One potential explanation for this pattern of results is that, when trying to remember the two line orientations, participants may have encoded the two distinct orientations as one "chunked" representation. Participants could have bound the two distinct orientation objects into one chunked representation, with both orientations bound together as an angle or clock hands, for example. Anecdotal evidence from post-experimental debriefing of our participants is consistent with this interpretation. Although the two Gabor orientations were presented separately in the lower left and right hemifields, many participants reported encoding the two as a bound object, (e.g., an angle by projecting the lines out to their intersecting point, like the hour and minute hands on an analog clock). Encoding the two objects as a bound object would change the way the relevant and irrelevant features are represented. If two stimuli retained in WM are bound or "chunked" into a single object, then it may not be possible to fully delete the no-longer-relevant item from WM following a retrocue. This might explain the pattern of results, which differs from paradigms with retrieval of more discrete (e.g., categorical) stimuli that cannot be as easily bound into a single object, such as a face paired with either a word or a direction of motion, as in Rose et al. (2016) (see also Fulvio & Postle, 2020). To test this account of the biases from the no-longer-relevant stimulus a second experiment was conducted.

Experiment 2

Experiment 2 used the same task design as Experiment 1. The only difference was that we explicitly instructed participants to bind or "chunk" the two orientations together on each trial by mentally connecting the line orientations into one bound object. We told participants to imagine the line orientations' point of intersection and think about the two orientations as an angle or hands of a clock. If the source of the bias from the no-longer-relevant item on recall of the target item that was observed in Experiment 1 was due to this binding at encoding, then the pattern of results should be similar for Experiment 2. If the pattern is not similar then the source of bias must be due to some other factor.

Method

The method for Experiment 2 was identical to Experiment 1 except that, during the practice, participants were read the following instruction: "In order to store the orientations of the two gratings, visualize them as the hands of a clock; project the lines out to their point of intersection and remember them as an angle like the hands of the clock. This might help you to remember them more easily." All other methods remained the same as Experiment 1, including the viewing distance and stimulus locations.

An a priori power analysis indicated that the minimum sample size needed to attain an effect size as large as the effect reported in Experiment 1 (effect size d = 0.61) was N = 31 with 95% power and α = .05.

Participants Thirty-three (M_{age} = 18.94 years (SD = 1.41), range = 18–26 years, 22 female) right-handed students with normal or corrected-to-normal vision participated in the experiment. Participants provided informed consent (IRB protocol 17-02-3629) and were remunerated with cash or course credit (US$15 or 1 credit/h).

Data quality checks As in Experiment 1, for the average accuracy analysis, in order to identify potential outliers in the data, for each recall condition per participant, the errors were first converted to z-scores and any z-score >3 or <−3 was removed from the data set. Because these responses were significant outliers, they likely reflect cases in which participants had no memory representation for the target item and resorted to guessing. Therefore, removing these responses before analysis enabled us to get a more accurate measure of memory performance. A total of 1.6% responses were removed (233 out of a total of 14,336 responses), and no more than seven trials were removed from any recall condition for an individual participant. The analysis of errors was conducted on the non-z-score converted data as the circular deviation of the recalled orientation from the target orientation in degrees. The *boxplot* function in R Studio was then used across all participants to determine any outliers in the dataset for each recall condition (defined as 1.5 times the interquartile range above or below the third or first quartiles, respectively). One participant was determined to be an outlier, and the error analyses were conducted on the remaining 32 participants' data.

For the mixture model analyses, all trials for the remaining participants were included (even the trials previously identified by the z-score analysis as outliers) because the models attempted to separate errors by different parameters, so were able to account for data points (e.g., swap errors) that may appear as outliers. For the nontarget bias analysis, we included all trials as in the average accuracy analysis (except those from the excluded outlier subject).

Nontarget bias analysis The analysis of the effect of the nontarget memory item on the recall of the target item was approached in the same way for Experiment 2 as in Experiment 1, but the difference between the experiments required a change in how we calculated the difference between the nontarget memory item and the response. The Gabor orientations used in these experiments are bidirectional (which means 0° and 180° are perceptually identical) rather than unidirectional (such as teardrops or lines with an arrowhead). In Experiment 1, because the two stimuli were assumed to be encoded independently and were bidirectional, the differences between the nontarget and the response orientations were calculated based on the smallest angular difference between them, so errors could not be greater than 90° (meaning, we assumed that the side of the response orientation that was closest to the stimulus was the side the participants used when making their response).

Because Experiment 2 instructed participants to bind the two stimuli together at the intersecting vertex, the Gabor orientations would have directional information associated with them; participants would have been maintaining and responding to angles that were sometimes obtuse (larger than 90°). Therefore, for Experiment 2 the difference between the nontarget and response orientations must be calculated based on the bound angle that the participants were instructed to attend to and maintain, not the smallest possible angle between the target and nontarget stimuli. Thus, the bias analysis was conducted with six bins to span the full 180° space from small to large differences between the orientations, rather than three bins to span 0–90° as was done in Experiment 1. As a result, some errors that would have been considered attractive in Experiment 1 were calculated to be repulsive in Experiment 2, and vice versa. Note that we recalculated the differences between the nontarget and the response orientations on each trial from Experiment 1 according to this scheme in order to assess the extent to which this affected the bias analyses. Doing so did not substantially alter the pattern of results or the main conclusions (see OSM Figs. 4 and 5).

Results

The distributions of recall errors relative to the to-be-remembered target orientation on recall-1, recall-2-stay, and recall-2-switch trials for all participants are shown in Fig. 6. The same series of analyses were conducted for Experiment 2 as in Experiment 1. Then, to determine the consequences of holding information in an unprioritized state when the information is a feature bound to another feature, we report

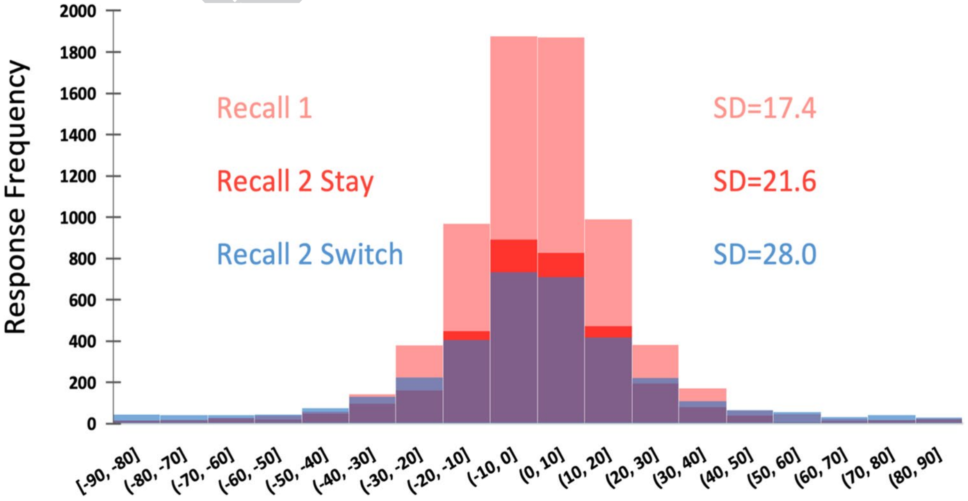

Fig. 6 Experiment 2: The frequency of recall errors and the standard deviation (SD, i.e., memory precision) in degrees relative to the target orientation for each condition (recall-1, recall-2-stay, and recall-2-switch) for all trials and all participants. Memory precision was similar for recall-1 and recall-2-stay trials, while recall-2-switch trials were less precise. Note that, by design, recall-1 had twice as many trials as recall-2-stay and recall-2-switch trials; also note the lack of any systematic bias to the left (negative degrees) or right (positive degrees) of the target orientation

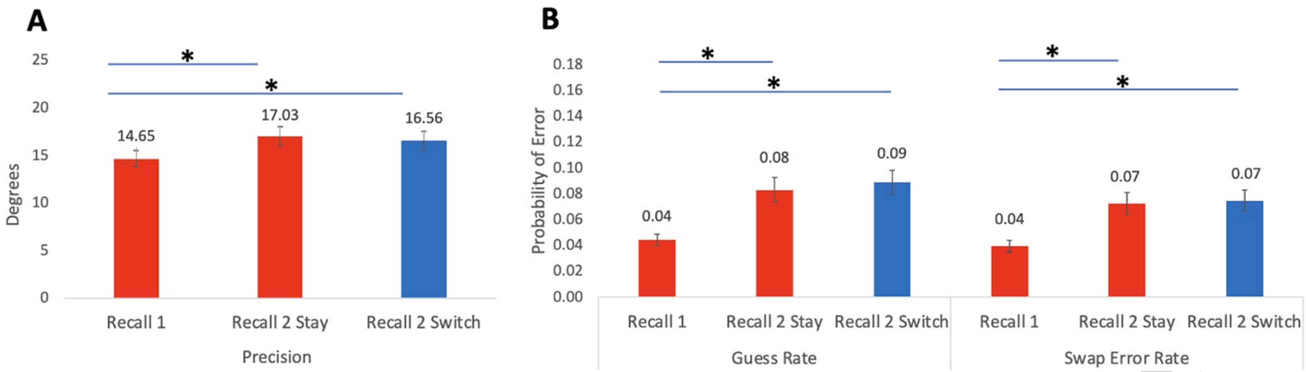

Fig. 7 Experiment 2: Average swap model parameter values. **A** The average precision parameter in degrees for the three recall conditions (lower values indicate better precision). Memory was less precise for both recall-2-switch and recall-2-stay trials compared to recall-1 trials. **B** The average guess rate and average swap error rate parameters for the three recall conditions. Both the average guess rate and average swap error rates were higher for both recall-2-stay and recall-2-switch trials compared to recall-1 trials. Error bars reflect ±1 standard error of the mean; $*p < 0.001$

the analyses that directly compared the results between Experiments 1 and 2. For all analyses, two-tailed, Bonferroni-corrected t-tests were used, unless stated otherwise

We first compared the average absolute value of recall error (in degrees) across the three recall conditions (recall-1, recall-2-stay, and recall-2-switch trials). As in Experiment 2, recall error was higher on recall-2-switch trials ($M = 19.3$, $SD = 7.34$) than on both recall-1 ($M = 12.4$, $SD = 4.42$; $t(31) = -7.86$, $p < .001$) and recall-2-stay trials ($M = 14.5$, $SD = 5.20$; $t(31) = -7.45$, $p < .001$). These results support our hypothesis that shifting a memory item into an unattended state weakened the fidelity of memory for that item compared to items maintained in an attended state. However, recall error was also higher on recall-2-stay than on recall-1 responses ($t(31) = -3.00$, $p < 0.01$), which contrasts with the result in Experiment 1. Next we conducted mixture model analyses on the data in order to better understand the source(s) of the differences.

Mixture modeling

Model preference We first compared the two mixture models to determine which of the models was a better fit to the data. We performed a Wilcoxon signed-rank test on the difference in the AIC values between the Standard Mixture Model and the Swap Model for all participants (as in Bays & Taylor, 2018). As in Experiment 1, the Swap Model was preferred over the Standard Mixture Model for all three recall conditions: recall-1 ($\Delta M = 10.81$, $p < .001$); recall-2-stay ($\Delta M = 5.57$, $p < .01$); and recall-2-switch ($\Delta M = 7.1$, $p < .01$).

Parameter differences To determine the effects of shifting attention between items, we compared the error parameters from the Swap Model across the three recall conditions (recall-1, recall-2-stay, and recall-2-switch). Three paired t-tests were performed for each parameter to compare all three recall conditions.

Precision Consistent with Experiment 1, there was a statistically significant increase in the precision parameter (indicating less precision) on recall-2-switch trials than recall-1 trials ($t(31) = -3.99$, $p < .01$). In contrast to Experiment 1, there was also a statistically significant increase in the precision parameter on recall-2-stay trials compared to recall-1 trials ($t(31) = -4.55$, $p < .01$), and there was no significant difference in the precision parameter between recall-2-stay trials and recall-2-switch trials ($t(31) = -1.41$, $p = 0.17$, Fig. 7A).

Guess rate Consistent with Experiment 1, the guess rate was higher for recall-2-switch trials than recall-1 trials ($t(31) = -6.03$, $p < .001$). In contrast to Experiment 1, the guess rate was higher for recall-2-stay trials than recall-1 trials ($t(31) = -4.78$, $p < .001$), and there was no significant difference between recall-2-stay and recall-2-switch trials ($t(31) = -1.44$, $p = 0.16$, Fig. 7B).

Swap error rate Consistent with Experiment 1, the estimated swap error rate was higher on recall-2-switch trials than recall-1 trials ($t(31) = -5.60$ and -4.48, $ps < .01$);[8] in contrast to Experiment 1, the estimated swap error rate was also higher on recall-2-stay trials than recall-1 trials ($t(31) = -4.48$, $ps < .01$), and the difference in the swap error rate between recall-2-stay and recall-2-switch trials was not significant ($t(31) = 0.61$, $p = 0.54$, Fig. 7B).

Thus, the main difference between Experiments 1 and 2 was poorer precision, guess, and swap error rates for recall-2-stay trials. To elucidate the source of the differences between the recall conditions we investigated the role that

---

[8] Note that, as in Experiment 1, these are one-tailed t-tests because of our a priori hypothesis that swap errors would increase on recall-2-switch trials.

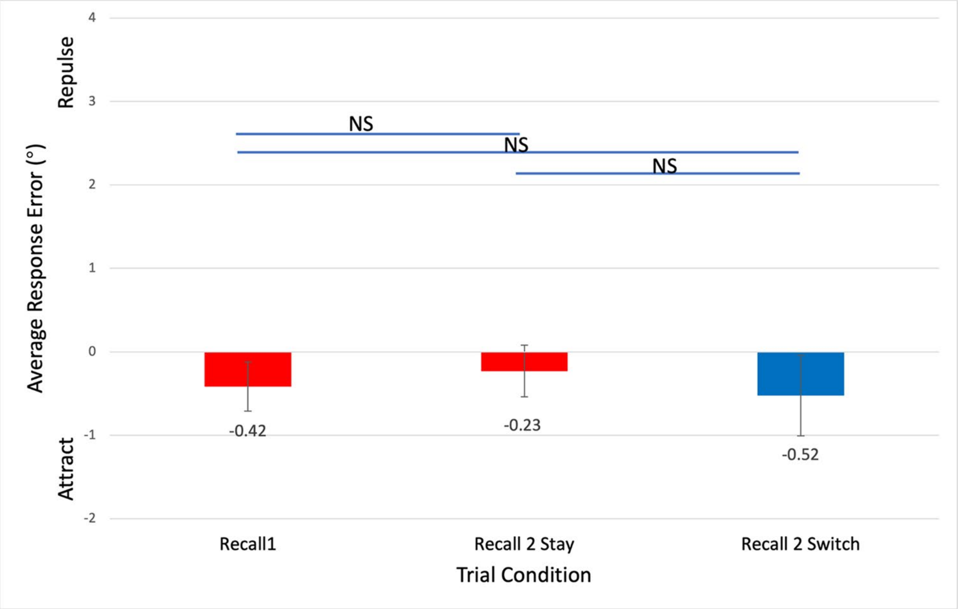

Fig. 8 Experiment 2: Average response error bias from the nontarget item for each trial condition. Responses were calculated based on whether errors were committed closer to (less than 0) or away from (greater than 0) the orientation of the nontarget item and averaged for each trial condition. All trials exhibited an attractive bias toward the nontarget item. Error bars reflect 1 SEM; NS indicates a Non-Significant difference at $p < .05$ (uncorrected)

the nontarget played in biasing response errors and how this bias changed across the recall conditions.

Nontarget bias analysis: repulsion and attraction effects The average response error relative to the nontarget for the three recall trial conditions is shown in Fig. 8. In contrast to Experiment 1, the average response error was not significantly different from zero for all three conditions ($t(31)$s < 1.44, $p > 0.16$), and there were no significant differences between the three conditions ($t(31)$s < 0.51, $ps > 0.19$).

However, the amount of bias varied as a function of the difference between the target and nontarget orientations. That is, the amount of bias that the uncued (nontarget) item had on recall of the cued (target) item depended on the similarity between the target and nontarget. Trials were binned around six orientation differences centered around relatively small (25°) to large (150°) differences between the target and nontarget orientations (see Fig. 9). There were significant main effects of condition ($F(2,62) = 9.754$, $p < 0.05$) and bin ($F(5,155) = 27.513$, $p < 0.001$), and there was a significant interaction between condition and bin ($F(2, 5,310) = 12.332$, $p < 0.001$).

For recall-1 responses, there was not significant bias from the nontarget, except on trials with large (~150°) differences between the target and nontarget (M = 5.94, p < .001). For both recall-2-stay and switch trials, there was significant bias that was either repulsive when the target-nontarget difference was just clockwise of the cardinal axes (0° and 90°, i.e., ~25° and ~100° bins, $ps < .05$) or attractive when the difference was just counterclockwise of the cardinal axes (90° and 180°, i.e., ~75° and ~150° bins, $ps < .001$, except for the 150° bin for recall-2-switch trials, Bonferroni corrected $p = .12$). When the target-nontarget difference was far from a cardinal axis (trials in the ~50° or ~125° bins) there was not significant bias for any recall condition (except for the 50° bin for recall-2-stay trials, $p < .05$). Recall-2-stay and recall-2-switch trials did not significantly differ from one another in the amount or type of bias (repulsion vs. attraction) except on trials with small (~25°) differences between the target and nontarget (M = 6.33 vs. 2.77, $p = .014$). In sum, as in Experiment 1, the amount of bias from the nontarget was larger on recall 2 than recall 1 responses, which contradicts the active-deletion hypothesis.

Are differences in precision, guess, and swap parameters due to bias? As in Experiment 1, an exploratory correlational analysis was done to see if the poorer precision, guess, and swap error parameters on recall-2 stay and switch trials that were observed (Fig. 7) were associated with the increase in repulsive bias that was observed on these trials (Fig. 9). Participants' precision, guess, and swap parameters on recall-1, recall-2-stay, and recall-2-switch trials were not correlated with their average amount of bias on those trials ($rs < .22$, $ps > .23$), indicating that the increase in precision, guess and swap errors was not associated with the increase in bias on recall-2-stay or switch trials.

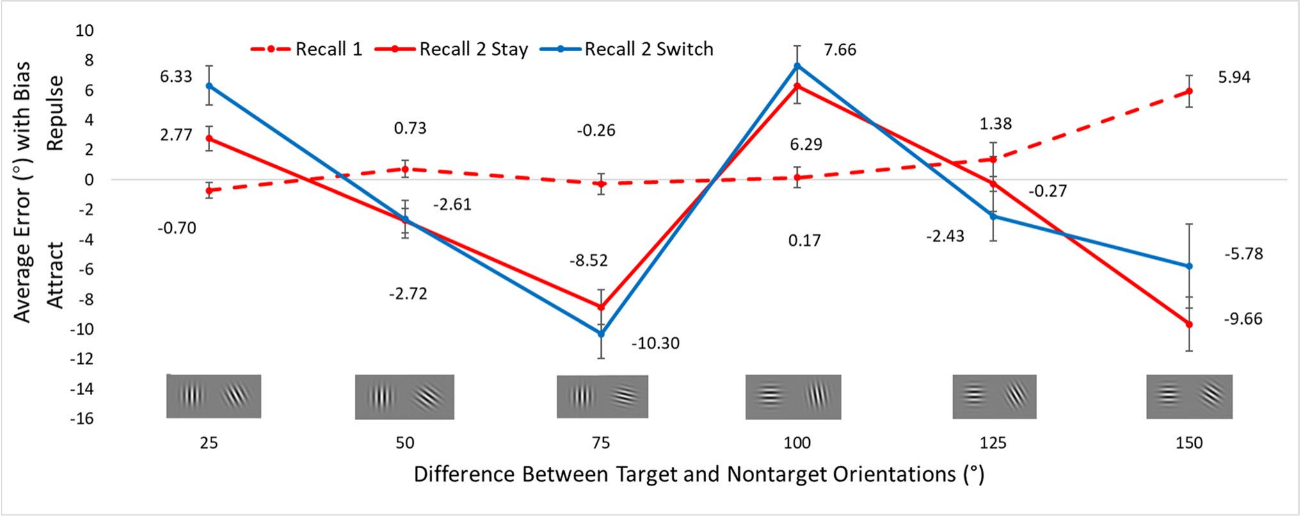

Fig. 9 Experiment 2: Average response error bias as a function of distance between orientation stimuli. Response errors were calculated as the degree of difference from the target orientation towards (negative) or away from (positive) the nontarget item, and the average response bias across participants was plotted for each recall condition (Note that Experiment 1 data was reanalyzed using the six-bin method from Experiment 2 in order to facilitate cross-experiment comparisons. See *Bias calculation* and OSM Fig. 5 for a more detailed explanation and rationale). Bins were created using trials in which the difference between the stimuli were ± 11.5° from 25°, 50°, 75°, 100°, 125°, and 150°, respectively. Error bars reflect ±1 standard error of the mean

Discussion

The second experiment was conducted to test if the source of the bias from the no-longer-relevant orientation on recall-2-switch trials in Experiment 1 was because participants had bound the two orientations together into one object, for example, as an angle or clock hands. If so, then the results of Experiment 2 should have been similar to those of Experiment 1. The effects of the encoding manipulation were assessed by comparing the results between Experiments 1 and 2 with between-group statistical tests. These comparisons showed that instructing participants to bind the orientations together had the following impacts:

The distributions of recall errors (SD) were reduced for recall 1 (from 22.5 to 17.97), recall-2-stay (from 22.7 to 19.62), and recall-2-switch trials (from 32.7 to 27.59) (compare Figs. 2 and 6). Chi-square tests showed that the reductions in the distribution of recall errors between Experiments 1 and 2 were significant for each recall condition (recall-1: χ^2 (17, N = 7,388) = 187.09, $p < .01$, recall-2-stay: χ^2 (17, N = 7,388) = 74.57, $p < .01$, recall-2-switch: χ^2 (17, N = 7,388) = 352.59, $p < .01$).

To see if mean recall error differed between the two experiments for each condition, independent-samples t-tests were conducted on the absolute difference in recall from the target orientation. These showed that recall was better for Experiment 2 than Experiment 1 for recall-1 ($t(63) = 2.04$, $p < 0.05$) and recall-2-switch trials ($t(63) = 1.98$, $p < 0.05$), but not recall-2-stay trials ($t(63) = 1.78$, $p > 0.05$).

The mixture model fits were compared with independent samples t-tests on the mean AIC values for each condition. These showed that the model fits for each recall condition did not differ between Experiments 1 and 2 (recall-1: $t(62) = 1.26$, $p > 0.05$, recall-2-stay: $t(62) = 0.78$, $p > 0.05$, recall-2-switch: $t(62) = 0.87$, $p > 0.05$). This indicates that the swap model was preferred over the standard mixture model to a similar degree in both experiments for each recall condition.

However, whereas the precision, guess rate, and swap error rate parameters were unchanged for recall-1 and recall-2-stay trials between Experiments 1 and 2, all three parameters improved for recall-2-switch trials (compare Figs. 3 and 7). For recall-2-switch trials, there was significantly lower (meaning better) precision ($t(62) = 2.18$, $p < 0.01$), and there was a reduction in both swap error ($t(62) = 2.43$, $p < 0.01$) and guess rates ($t(62) = 3.29$, $p < 0.01$) in Experiment 2 compared to Experiment 1.[9] In contrast, for recall-1 and recall-2-stay trials, there were no significant differences between the two experiments for any of the parameter estimates [recall-1: guess rate ($t(62) = 1.73$, $p > 0.05$), swap error rate ($t(62) = 1.20$, $p > 0.05$, precision ($t(62) = 0.10$, $p > 0.05$)); recall-2-stay: guess rate ($t(62) = -0.31$, $p > 0.05$), swap error rate ($t(62) = -1.18$, $p > 0.05$), precision ($t(62) = -1.63$, $p > 0.05$)].

---

[9] These results, particularly the decrease in swap error rates, suggest that the participants did indeed follow instructions and bind the two orientations as one object, resulting in a lower frequency of binding errors between each orientation and its location.

With regard to the analyses of the bias from the nontarget orientation on recall of the target orientation, the overall bias, averaged across trials with small to large differences between the target and nontarget orientations, was no longer significant (compare Figs. 4 and 8). However, the amount and type of bias (repulsive vs. attractive) differed on trials as a function of the degree of difference between the target and nontarget orientations. Bias changed from having more bias for recall-2-switch than both recall-2-stay and recall-1 in Experiment 1 (especially for trials with similar target-nontarget differences), to having more bias for both recall-2-switch and recall-2-stay than recall-1 in Experiment 2. Additionally, the nature of bias in Experiment 2 appeared to switch between repulsion and attraction depending on how close the target-nontarget difference was to the cardinal cartesian axes (0°, 90°, 180°) (compare Figs. 5 and 9).

Formally testing for differences in the bias effects between the two experiments was complicated by the fact that, in Experiment 1, the target-nontarget difference between the non-directional orientations spanned from 0° to 90°, with average bias measured within three orientation bins (25°, 50°, and 75° ± 11.5°). In Experiment 2, because participants were to bind the orientations together as segments of an angle, the target-nontarget difference in orientations spanned from 0° to 180°, so average bias had to be measured within six orientation bins (25°, 50°, 75°, 100°, 125°, and 150° ± 11.5°). We first compared bias between the experiments for the 25°, 50°, and 75° bins for each recall condition with independent samples t-tests. For trials with relatively small target-nontarget differences (25° ± 11.5°), there was no difference in bias between Experiments 1 and 2 for each recall condition ($ts(63) < 2.78$, $ps > 0.064$). For trials with medium (50° ± 11.5°) target-nontarget differences, recall-1 and recall-2-stay did not show a significant difference in bias between Experiments 1 and 2 [recall-1: $t(63) = -1.13$, $p = 1.000$; recall-2-stay: $t(63) = 2.10$, $p = 0.357$]; recall-2-switch showed repulsion in Experiment 1, but attraction in Experiment 2 ($t(63) = 3.78$, $p = 0.003$). For trials with large (75 ± 11.5°) target-nontarget differences, bias was not different between the experiments for recall-1 ($t(63) = 0.35$, $p = 1.000$). For both recall-2-stay and recall-2-switch trials there was more attractive bias in Experiment 2 than in Experiment 1 [recall-2-stay: $t(63) = 6.49$, $p < 0.001$; recall-2-switch: $t(63) = 6.07$, $p < 10.001$].[10]

To summarize the results from Experiment 2 and their comparison to those from Experiment 1, there was still the greatest bias from the nontarget orientation on recall of the target (repulsive or attractive) when it was no longer relevant (on recall-2-stay and switch trials) even though, according to the active-deletion hypothesis, the nontarget should have been removed from WM and exerted less bias on recall-2 than recall-1 trials. This bias was observed on trials when the target-nontarget difference was just clockwise or counterclockwise of the cardinal axes. When the target-nontarget difference was far from the cardinal axes, recall showed no bias from the nontarget orientation. While differences in the data between Experiment 1 and Experiment 2 suggest that binding the two stimuli was not the only source of bias in the Experiment 1 data, the pattern of results from Experiment 2 nonetheless provides further evidence for the need to revise the active-deletion hypothesis.

General discussion

The purpose of this study was to test the active-deletion hypothesis – that no-longer-relevant/nontarget information is actively deleted from WM so that it does not interfere with memory for relevant/target information in WM. The main finding that was observed is that the strongest amount of bias on recall of a target orientation maintained in WM was from a no-longer-relevant/nontarget orientation that, according to the active-deletion hypothesis, should have been removed from WM. Following Experiment 1, we hypothesized that this bias was present because participants may have been binding the two orientations together into one object so that, when the nontarget orientation was cued as no longer relevant on the trial, it may not have been possible to actively delete the nontarget orientation from WM. To test this hypothesis, we conducted a second experiment in which we explicitly instructed participants to bind the two orientations together and think of them as two line segments in an angle as in the hour and minute hands of an analog clock. Once again, the amount of bias on WM recall was strongest from the nontarget item that, according to the active-deletion hypothesis, should have been deleted. This strengthens confidence in the main conclusion from Experiment 1 – that no-longer-relevant items (orientations), that should have been deleted from visual WM, were not removed from WM – they continued to exert bias on WM performance even when they become irrelevant for ongoing cognition. These results call for a revision to the active-deletion hypothesis and models of WM.

Relation to prior research

One previous study utilized a task design with some similarities (and some important differences) and also showed a repulsive bias on recall when a no-longer-relevant item should have been deleted from WM (Bae & Luck, 2017).

---

[10] For completeness, the bias analysis for Experiment 1 was recalculated using six bins to facilitate comparison between Experiments 1 and 2 (see OSM Fig. 5). As in the three-bin analysis, there were no differences in bias between recall-1 and recall-2-stay trials for any bin, and there was greater bias on recall-2-switch trials, especially for the 50°and 125° bins.

Bae and Luck (2017) did not discuss the implications that this finding has for the active-deletion hypothesis. As discussed below, that a repulsive bias from a no-longer-relevant item was found in both of these two independent studies from different labs using tasks with some important differences in methods is compelling and strengthens confidence in the robustness of this phenomenon.

Our results converge with those of Bae and Luck (2017) despite some important differences between the experimental paradigms. In their study, directional orientations ("teardrops") were presented sequentially, overlapping at central fixation, whereas in the present study Gabor patches were presented simultaneously, separated by approximately 22.3° of visual angle from one another in the lower left and right hemifields. These are not trivial methodological differences. There was considerably more overlap in the cortical areas that processed the visual stimuli in their experiment than ours, so it was plausible that there would be stronger modulation of local cortical circuits (via lateral inhibition) that repulsed the memory representations of the stimuli in their experiment than ours (Johnson et al., 2009).

Also, sequentially presenting the stimuli at the same location could have resulted in substantial bias from lateral inhibition because the memory representation for the second item could have included relative information (e.g., X° clockwise/counterclockwise from the first stimulus). Furthermore, in their sequential report paradigm, both items were always tested, the order of recall was determined by the first retrocue, and the second item was recalled immediately following the first item. The short interval between recalling the first and second item could have caused more bias from the first item on the second item than with our paradigm, perhaps because there was not enough time for participants to actively delete the no-longer-relevant orientation from WM (Oberauer, 2018).

In our paradigm, the item to be recalled second was unknown until the second cue appeared, which was several seconds after recalling the first item (a much longer period than in their paradigm). Nevertheless, the data from both studies showed strikingly similar evidence that items which (according to the active-deletion hypothesis) should have been deleted from WM continued to bias retrieval of a target item in WM; this diverges from evidence supporting the notion that no-longer-relevant items do not influence retrieval of target items in WM (i.e., Fulvio & Postle, 2020; Rose et al., 2016). The results reported here and by Bae and Luck (2017) show that this clearly was not the case.

What might explain the source and direction of biases on WM?

What might be the source of the biases that were observed in this study? An alternative to the binding/chunking account is that the bias seen on recall-2-switch trials in Experiment 1 could be from the recalled orientation for the recall 1 response on those trials, as opposed to the memory representation that was encoded, cued, maintained, and retrieved for recall 1. Note that the active-deletion hypothesis suggests that no-longer-relevant information should be deleted from memory, so any memory of the recall 1 response should also have been actively deleted from WM so that this irrelevant information did not interfere with WM for the second-cued, target item. Nonetheless, a future study with this paradigm that includes trials in which the item cued first is not tested would shed light on the extent to which memory of the response, rather than memory of the stimulus, drives the observed biases. It is noteworthy that at least two studies have included such trials and ruled out this possibility (see Dagry et al., 2017 and Lintz & Johnson, 2021).[11]

There is a long history of related research on proactive interference and "serial dependence" effects showing that a response to a target item on the current trial is biased from previous responses to targets on previous trials. There is a robust literature on such effects that span the perception, attention, and memory domains, so review of these literatures is beyond the scope of the current study (e.g., Bliss et al., 2017; Fischer & Whitney, 2014; for reviews, see Kiyonaga et al., 2017; Lorenc et al., 2021). If recall on this task was biased by the previous response, then recall 1 responses in the current study should also have been biased by the response from the previous trial. To test this hypothesis, a supplemental analysis was conducted by recalculating the bias on recall 1 errors based on the previously recalled response on recall 2 from the previous trial. The amount of bias on recall 1 responses from the recall 2 response on the previous trial was not as large as the bias observed on recall-2-switch trials (see OSM Fig. 6). Therefore, the amount of bias from the no-longer-relevant item within the same trial was stronger than any bias seen from the previously recalled item on the previous trial (i.e., proactive interference or serial dependence). Future studies that are designed to directly compare the size and nature of bias effects from proactive interference and serial dependence from interfering items encoded, attended, or retrieved on previous trials

---

[11] Note that the feedback following recall 1 could result in corrections to the errors made, so memory may have also been influenced by feedback, especially for recall-2 trials. To assess the extent to which this affected recall, a supplemental analysis was done to compare recall-2 trials following green, yellow, or red feedback on recall-1 responses. This showed that the effects of feedback on recall were minimal and did not substantially alter interpretation of the observed effects (see OSM Figs. 2 and 3). Nonetheless, future studies should manipulate whether feedback is provided so that the independent effect of feedback on memory can be assessed. We thank Evan Lintz for this suggestion.

versus within the same trial are needed to elucidate the source of these interesting bias effects.

What determines when there will be repulsive or attractive bias from nontargets? Although addressing this question is beyond the scope of the current study, it is an interesting question that emerges from the results. Some recent research reported that WM for an item on the current trial was attracted to the memory item from the previous trial, but the direction of this bias flipped to be a repulsive bias three trials later (Fritsche et al., 2020). The authors interpreted this to be due to influence from both Bayesian priors and efficient encoding similar to perceptual adaptation. Note that studies that have shown attractive or repulsive bias often involve distractors that are irrelevant to task performance – that is, the distractors were never maintained in WM (e.g., Mallett et al., 2020). Such studies do not provide the most direct tests of the active-deletion hypothesis. The same is true of studies involving task situations in which there is an insufficient amount of time to actively remove a no-longer-relevant item from WM (Golomb, 2015).

In at least one study that can provide more of a direct test of the active-deletion hypothesis, Czoschke et al. (2019) suggested that attractive bias is seen from distractors occurring across trials whereas repulsive bias is seen from distractors occurring within trials. Chunharas et al. (2022) suggest that bias is attractive when the number of items to remember is close to a participant's WM capacity, but repulsive for smaller, sub-span set sizes, especially for longer delays. Shan and Postle (2022) suggest that whether there is attraction repulsion depends on whether a no-longer relevant item was passively or actively removed from WM. Using a clever design similar to our own, but with a distractor that appeared in a location that did or did not overlap with one of the stimuli, they found that an irrelevant memory item exerted attractive bias on recall in the no-overlap condition, but an (unexpected) repulsive bias in the overlap condition.

Here we showed, with only two simple features to remember over relatively long delays (compared to most visual WM paradigms, especially for recall-2 trials), that the amount and direction of bias (repulsion or attraction) depended on the nature of encoding (whether the features were bound into a single object), the angular difference between the two orientations, and its proximity to cardinal axes. In sum, the attraction/repulsion literature across perception, attention, and WM studies is decidedly mixed. A clarifying account that spans these domains is needed. Nonetheless, our results add interesting data showing further dynamic, contextual variability of the phenomena to this growing body of research.

Regardless of the direction of bias from no-longer-relevant distractors or the precise mechanisms that cause such bias (which are not entirely clear yet), the most important take home point is that evidence of such biases are inconsistent with the active-deletion hypothesis. So, this mechanism, which is hypothesized to help control the contents of WM by prioritizing maintenance of target information and resolve interference from nontarget information, does not appear to be used in all circumstances. Clarifying the exact source of such differences between studies which suggest that active deletion is or is not used is an important direction for future research on the dynamics of WM. Doing so will help researchers pin down why certain items persist in WM when others do not.

At present, the results of this study may be seen to support at least some of the conclusions drawn by Oberauer (2018) and the SOB-CS model. The data suggest that, even when low level visual stimuli are used as memoranda (rather than words), the simultaneous stimulus presentation and binding to each stimulus's spatial context involved sufficient processing (perhaps via chunking or the persistence of a previously retrieved representation). This may have prevented the no-longer-relevant item from being removed from WM in the 2.5 s between the second cue and the probe. Further exploring what level of processing of the stimuli is required to prevent active-deletion could shed light on possible mechanisms for this removal process and help researchers gain insight into when and why active-deletion occurs.

A limitation of this study is that, due to the COVID-19 pandemic, we were unable to use neuroimaging or neurostimulation methods to observe or modulate the activation status of items held in WM (as in Rose et al., 2016). Having participants perform a visual WM double-retrocue task with concurrent neuroimaging and neurostimulation, and associating neural data with potential biases from irrelevant items, could help reveal the nature of the representations that are retained in WM, including their activation state and the extent to which target and irrelevant features may be chunked or bias one another. For example, Bae and Luck (2019) were able to decode the no-longer-relevant orientation that was recalled on a previous trial from the EEG signals evoked by recall of a target orientation on the current trial. Such analyses can be used to track the activation state and the influence of WM items as they transition from relevant to no-longer-relevant, deleted states (see also Lorenc et al., 2020).

Additionally, future research would do well to assess the independent contributions of the effects of attentional cueing/prioritization separately from the act of recalling an initial target item on the nontarget biases that were observed on recall 2 trials. It will be important for future research to elucidate the source of biases from no-longer-relevant/nontarget items on recall of target items, and whether such interactions among items in WM arise during encoding, maintenance, or retrieval. Analyzing bias from no-longer-relevant items, and understanding how it interacts with different prioritization states, should help researchers elucidate the nature of WM representations and how they are influenced by other items in memory.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.3758/s13414-023-02724-2.

Acknowledgements This work was supported by the National Science Foundation CAREER Grant 1848440 awarded to N.S.R and from the University of Notre Dame William P. and Hazel B. Collegiate Chair endowment awarded to N.S.R. The design, analyses, and results of Experiment 1 were reported as part of J.P.R.'s Senior Thesis and his presentation in the 2021 Virtual Working Memory Symposium. Experiment 2 was designed by N.S.R. and conducted by C.X. We thank Rosanne Rademaker for some of the experiment code, Steve Emrich and Tim Brady for helpful advice on the modeling, and Jori Waner, Isaiah Metcalf, Cheick Diallo, Yuki Yoshioka, Lauren Crowe, and Bernadette Widjaja for contributing to data collection and analysis for Experiment 1, and Lauren Crowe, Savannah Gregory, Jo'Vette Hawkins, and Luke Borman for contributing to data collection for Experiment 2. The Data are publicly available at CurateND: https://doi.org/10.7274/r0-mq7c-7m28.

Declarations

Conflicts of interest The authors have no potential conflicts of interest to report.

References

Abrahamyan, A., Clifford, C. W. G., Ruzzoli, M., Phillips, D., Arabzadeh, E., & Harris, J. A. (2011). Accurate and Rapid Estimation of Phosphene Thresholds (REPT). *PLoS One, 6*(7), e22342.

Baddeley, A. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology, 63*, 1–29.

Bae, G. Y., & Luck, S. J. (2017). Interactions between visual working memory representations. *Attention, Perception, & Psychophysics, 79*(8), 2376–2395.

Bae, G. Y., & Luck, S. J. (2019). Reactivation of Previous Experiences in a Working Memory Task. *Psychological Science, 30*(4), 587–595.

Bays, P. M., & Taylor, R. (2018). A neural model of retrospective attention in visual working memory. *Cognitive Psychology, 100*, 43–52.

Bays, P. M., Catalao, R. F. G., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision, 9*(10), 7.

Bliss, D. P., Sun, J. J., & D'Esposito, M. (2017). Serial dependence is absent at the time of perception but increases in visual working memory. *Scientific Reports, 7*(1), 14739.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10*, 433–436.

Chunharas, C., Rademaker, R. L., Brady, T. F., & Serences, J. T. (2022). An adaptive perspective on visual working memory distortions. *Journal of Experimental Psychology: General*. https://doi.org/10.1037/xge0001191

Czoschke, S., Fischer, C., Beitner, J., Kaiser, J., & Bledowski, C. (2019). Two types of serial dependence in visual working memory. *British Journal of Psychology, 110*(2), 256–267.

Dagry, I., & Barrouillet, P. (2017). The fate of distractors in working memory: No evidence for their active removal. *Cognition, 169*, 129–138.

Dagry, I., Vergauwe, E., & Barrouillet, P. (2017). Cleaning working memory: The fate of distractors. *Journal of Memory and Language, 92*, 327–342.

Fischer, J., & Whitney, D. (2014). Serial dependence in visual perception. *Nature Neuroscience, 17*, 738–743.

Fritsche, M., Spaak, E., & de Lange, F. P. (2020). A Bayesian and efficient observer model explains concurrent attractive and repulsive history biases in visual perception. *Elife, 9*, e55389. https://doi.org/10.7554/eLife.55389.sa2

Fulvio, J. M., & Postle, B. R. (2020). Cognitive Control, Not Time, Determines the Status of Items in Working Memory. *Journal of Cognition, 3*(1), 1–8.

Golomb, J. D. (2015). Divided spatial attention and feature-mixing errors. *Attention, Perception, & Psychophysics, 77*(8), 2562–2569.

Hasher, L., Lustig, C., & Zacks, R. (2007). Inhibitory Mechanisms and the Control of Attention. In: *Variation in Working Memory*. 227-249.

Johnson, J. S., Spencer, J. P., Luck, S. J., & Schöner, G. (2009). A dynamic neural field model of visual working memory and change detection. *Psychological Science, 20*, 568–577.

Kiyonaga, A., & Egner, T. (2016). Center-surround inhibition in working memory. *Current Biology, 26*(1), 64–68.

Kiyonaga, A., Scimeca, J. M., Bliss, D. P., & Whitney, D. (2017). Serial dependence across perception, attention, and memory. *Trends in Cognitive Sciences, 21*(7), 493–497.

Kleiner M, Brainard D, Pelli D, 2007, "What's new in Psychtoolbox-3?" Perception 36 ECVP Abstract Supplement.

LaRocque, J. J., Eichenbaum, A. S., Starrett, M. J., Rose, N. S., Emrich, S. M., & Postle, B. R. (2015). The short- and long-term fates of memory items retained outside the focus of attention. *Memory and Cognition, 43*(3), 453–468.

Lewis-Peacock, J. A., Kessler, Y., & Oberauer, K. (2018). The removal of information from working memory. *Annals of the New York Academy of Sciences, 1424*(1), 33–44.

Lilienthal, L., Rose, N. S., Tamez, E., Myerson, J., & Hale, S. (2015). Individuals with low working memory spans show greater interference from irrelevant information because of poor source monitoring, not greater activation. *Memory & Cognition, 43*(3), 357–366. https://doi.org/10.3758/s13421-014-0465-3

Lintz, E. N., & Johnson, M. R. (2021). Refreshing and removing items in working memory: Different approaches to equivalent processes? *Cognition, 211*, 104655.

Lorenc, E. S., Vandenbroucke, A. R. E., Nee, D. E., de Lange, F. P., & D'Esposito, M. (2020). Dissociable neural mechanisms underlie currently-relevant, future-relevant, and discarded working memory representations. *Scientific Reports, 10*, 11195.

Lorenc, E. S., Mallett, R., & Lewis-Peacock, J. A. (2021). Distraction in visual working memory: Resistance is not futile. *Trends in Cognitive Sciences, 25*(3), 228–239.

Mallett, R., Mummaneni, A., & Lewis-Peacock, J. A. (2020). Distraction biases working memory for faces. *Psychonomic Bulletin and Review, 27*(2), 350–356.

Oberauer, K. (2009). Design for a working memory. *Psychology of Learning and Motivation, 51*, 45–100. https://doi.org/10.1016/S0079-7421(09)51002-X

Oberauer, K. (2018). Removal of irrelevant information from working memory: Sometimes fast, sometimes slow, and sometimes not at all. *Annals of the New York Academy of Sciences, 1424*(1), 239–255.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*, 437–442.

Peters, B., Rahm, B., Kaiser, J., & Bledowski, C. (2019). Differential trajectories of memory quality and guessing across sequential reports from working memory. *Journal of Vision, 19*(7), 3–3.

Rademaker, R. L., Van De Ven, V. G., Tong, F., & Sack, A. T. (2017). The impact of early visual cortex transcranial magnetic stimulation on visual working memory precision and guess rate. *PLoS One, 12*(4), e0175230.

Rose, N. S. (2020). The dynamic processing model of working memory. *Current Directions in Psychological Science, 29*(4), 378–387.

Rose, N. S., LaRocque, J. J., Riggall, A. C., Gosseries, O., Starrett, M. J., Meyering, E. E., & Postle, B. R. (2016). Reactivation of latent working memories with transcranial magnetic stimulation. *Science, 354*(6316), 1136–1139.

Sadil, P., Cowell, R., & Huber, D. E. (2021). The push-pull of serial dependence effects: Attraction to the prior response and repulsion from the prior stimulus. https://doi.org/10.31234/osf.io/f52yz

Schneegans, S., & Bays, P. M. (2017). Restoration of fMRI Decodability Does Not Imply Latent Working Memory States. *Journal of Cognitive Neuroscience, 29*(12), 1977–1994.

Shan, J., & Postle, B. R. (2022). The influence of active removal from working memory on serial dependence. *Journal of Cognition, 5*(1), 31.

Silvanto, J. (2017). Working Memory Maintenance: Sustained Firing or Synaptic Mechanisms? In: *Trends in Cognitive Sciences* (Vol. 21, Issue 3, pp. 152–154). Elsevier Ltd.

Souza, A. S., & Oberauer, K. (2016). In search of the focus of attention in working memory: 13 years of the retro-cue effect. *Attention, Perception, & Psychophysics, 78*(7), 1839–1860.

Souza, A. S., Rerko, L., & Oberauer, K. (2016). Getting more from visual working memory: Retro-cues enhance retrieval and protect from visual interference. *Journal of Experimental Psychology: Human Perception and Performance, 42*(6), 890–910.

Stokes, M. G., Muhle-Karbe, P. S., & Myers, N. E. (2020). Theoretical distinction between functional states in working memory and their corresponding neural states. *Visual Cognition, 28*(5–8), 420–432.

Suchow, J. W., Brady, T. F., Fougnie, D., & Alvarez, G. A. (2013). Modeling visual working memory with the MemToolbox. *Journal of Vision, 13*(10), 9.

Wallis, G., Stokes, M., Cousijn, H., Woolrich, M., & Nobre, A. C. (2015). Frontoparietal and cingulo-opercular networks play dissociable roles in control of working memory. *Journal of Cognitive Neuroscience, 27*(10), 2019–2034.

Wildegger, T., Myers, N. E., Humphreys, G., & Nobre, A. C. (2015). Supraliminal but not subliminal distracters bias working memory recall. *Journal of Experimental Psychology: Human Perception and Performance, 41*(3), 826–839.

Wolff, M. J., Jochim, J., Akyürek, E. G., & Stokes, M. G. (2017). Dynamic hidden states underlying working-memory-guided behavior. *Nature Neuroscience, 20*(6), 864–871.

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature, 453*(7192), 233–235.