Are Irrelevant Items Actively Deleted from Visual Working Memory?: No Evidence from Repulsion Effects in Dual-Retrocue Tasks

Joshua P. Rhilinger & Nathan S. Rose

University of Notre Dame

Address Correspondence to:

Nathan S. Rose, Ph.D.

390 Corbett Family Hall

University of Notre Dame

Notre Dame, IN 46556

+1 574.631.5473

nrose1@nd.edu

Abstract

In addition to processes associated with maintaining, manipulating, and updating to-be-remembered information for ongoing cognition, some theories suggest that working memory (WM) also involves the active deletion of irrelevant information, including items that were retained in WM, but are no longer relevant for ongoing cognition. Considerable evidence provides support for an active deletion mechanism, particularly for categorical representations (Rose et al., 2016; Fulvio & Postle, 2020; but see Bae & Luck, 2017 for contradictory evidence with line orientations). On each trial of the current task, healthy young adults maintained two line orientations in visual WM, switched attention to maintaining and recalling the orientation cued first, and then switched to recall the item cued second, at which point the uncued orientation was no-longer-relevant on the trial. The results showed that the no-longer-relevant items exerted the strongest "repulsive" bias on participants' recall of to-be-remembered items, directly contradicting the active deletion hypothesis. We suggest that visual WM binds features like line orientations into ensemble representations, and an irrelevant feature of a bound object cannot be actively deleted--it biases recall of the target feature via repulsion. Models of WM will need to be updated to explain this dynamic phenomenon.

Introduction

Working memory (WM) is used to maintain and manipulate multiple items in mind over short periods of time while using those memory items to accomplish a goal (Baddeley, 2012). WM is implicated in most forms of higher-order cognition (Conway et al., 2012), so a great deal of research has attempted to elucidate the neurocognitive mechanisms that support WM (Oberauer et al., 2018). WM is tightly linked to the perception, attention, long-term memory, and action systems because of the need to encode, attend to, retrieve, and act on task-relevant items retained in WM (Cowan, 1998). Many theorists propose that, in addition to actively retaining goal-relevant information, WM also involves actively deleting information that is no longer relevant through a process that is distinct from forgetting[1] (Hasher, Lustig, & Zacks, 2007; Lewis-Peacock, Kessler, & Oberauer, 2018; Oberauer, 2018). However, the existence and nature of this mechanism, and the extent to which it is under cognitive control, is under debate. This paper attempts to address this issue.

Measuring the prioritization and deletion of items in WM with retrocue tasks

Because performance worsens as the number of items held in WM increases, being cued to which item(s) will be tested is typically beneficial (Souza & Oberauer, 2016). Cueing provides advantages even when presented after items have been encoded (retrocues). In conditions with versus without retrocues, numerous studies have shown differences in behavior (accuracy, response times) and brain activity (EEG/ERP, MEG, fMRI) associated with cued versus uncued items (for a meta-analysis, see Wallis et al., 2015).

[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 is needed.

Many theorists have concluded that these differences arise when internal attention selects and protects cued/relevant items against interference from other items in WM (Duarte et al., 2013; Gunseli et al., 2015; Heuer & Schubö, 2016; Makovski, Sussman, & Jiang, 2008; Souza, Rerko, & Oberauer, 2016). Other theorists have posited that cueing 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 people selectively delete irrelevant items, or both (Lintz & Johnson, 2021). This is particularly unclear with single retrocue tasks because only cued items are tested. Tasks with multiple retrocues are more revealing because memory for the initially uncued/irrelevant item can be assessed. If switching attention away from uncued item(s) to attend to a cued item involves actively deleting one or more no-longer-relevant-items from WM, then the deleted items should not affect WM performance.

Using a two-item double-retrocue task, 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 (a face, word, or direction of motion) were initially presented and retained in WM, the category of both items could be decoded from fMRI or EEG. Following the first retrocue, neural representation of the uncued item dropped to baseline as if it were no longer "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, and this caused an increase in false alarms to recognition probes matching the uncued item. This suggested that UMIs were retained in an "activity-silent" manner via short-term synaptic plasticity mechanisms (Silvanto, 2017; Rose, 2020). Critically, following the second retrocue, which indicated that the uncued item was no longer relevant on the trial, TMS could no longer reactivate and increase false alarm rates for the uncued (no-longer-relevant) item. This suggested that the retrocues served to update WM and that items in WM were deleted when they were no longer needed to carry out the task (for replications, see Wolff et al., 2017; Fulvio & Postle, 2020).

Related research has also shown evidence of an active-deletion process that removes items cued as no longer relevant for WM storage (for review, see Lewis-Peacock et al., 2018). Despite evidence supporting the existence of a mechanism that helps to update the contents of WM to retain task-relevant items and limit interference from other items, it does not seem to always be utilized.

Repulsion Effects

Bae and Luck (2017) used a WM task similar to a double-retrocue task that required participants to recall two line orientations, which were presented at central fixation, in succession. The retrocue indicated which orientation to report first, and then which orientation to report second on each trial. Recall of both the first and second orientations were systematically biased away from ("repulsion") the other nontarget orientation.[2] This repulsion bias when recalling the second orientation is striking because the other orientation was no longer relevant to the task at the time of recall and should have been cleared from WM. That the no-longer-relevant item still biased recall of the second orientation contradicts the active-deletion hypothesis. In fact, bias from the nontarget was greater for the second than the first item recalled.

The Present Study

We attempted to replicate and extend this study by comparing how uncued (but potentially relevant) and "deleted" (no-longer-relevant) items bias recall of target items in WM.

[2] This repulsion phenomenon has also been shown to bias WM for faces (Mallett, Mummaneni, and Lewis-Peackock, 2020), motion (Czoschke et al., 2019), and color (Golomb, 2015).

We designed a double-retrocue task and analysis plan that allowed the measurement of memory fidelity of items held in attended, unattended, or the putative "deleted" state, as well as the relative contributions of distinct sources of influence on their recall. Specifically, computational modeling separated memory errors into three categories: 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 a response toward the memory item that was not cued for recall on that trial; also known as a binding error) (Peters et al., 2019).

We predicted that, relative to recall of the item cued first (and recall of the same item on "stay" trials), switching to recall the initially uncued item would be associated with worse precision, a higher guess rate, and a higher swap error rate. With regards to potential biases on recall, if the no-longer-relevant item is deleted from WM, as is suggested by the active-deletion hypothesis, then there should only be bias from the uncued, nontarget item when recalling the first item--not the second item when the nontarget is no longer relevant on the trial. That is, we should fail to replicate Bae and Luck (2017). If, however, bias from no-longer-relevant items is observed, this would call into question the generalizability of the putative active-deletion mechanism, at least for WM precision of low-level visual features like line orientations.

Methods

Subjects

A total of 41 subjects (14 Male, average age = 19.21; 27 Female, average age = 19.04) were recruited from the University of Notre Dame community to participate in the experiment. All subjects were between the ages of 18 and 35 and reported being right handed and having both normal color vision and normal or corrected-to-normal visual acuity. Subjects provided informed consent in accordance with the Notre Dame IRB protocols and were remunerated with cash ($15/hour) or course credit (1 credit/hour) through the University's SONA system. All blocks from participants in the sham rTMS condition (6 participants, approximately 46 blocks total) were included since these subjects did not undergo rTMS treatment; the rest of the subjects participated in a behavioral-only version of the experiment. Data from 6 subjects were not included (2 withdrew following the phosphene localization procedure, 1 was screened but never scheduled before COVID, data from the other 3 were lost due to technical errors), leaving a final total of 35 subjects whose data were analyzed.

Materials

Subjects were seated approximately 37 cm (n=19) or 57 cm (n=16) away from a 24-in ASUS computer monitor with 1920 x 1080 resolution and a 60 Hz refresh rate.[3] The task and stimuli were generated and run in MATLAB (R2014a, MathWorks, Natick, MA) using the Psychophysics Toolbox Version 3 (Brainard, 1997; Pelli, 1997; Kleiner, Brainard, & Pelli, 2007; http://psychtoolbox.org/). Responses were given on the T9 number pad of a standard QWERTY keyboard.

WM Task

Stimulus details

Central fixation was identified by a white circle with an outer radius of ¼ pixel and an inner radius of ⅛ pixel. The experimental stimuli consisted of two sine-wave gratings (i.e. Gabor

[3] The difference is due to a necessary change in the experiment room setup; note that the degree of visual angle at which stimuli were presented varied between subjects by design, see *Stimulus Details*, and did not interact with the effects of interest, see *Data Quality Checks*.

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 7 distinct orientation bins with centers of 13°, 39°, 65°, 91°, 117°, 143°, and 169°. For a given trial, orientations were selected pseudo-randomly from these bins with a jitter of ±5° (Samaha, Sprague, & Postle, 2016), and the two stimuli in a given trial varied by more than 10°.

Location of stimulus presentation was matched to phosphene localizations acquired from subjects in an ongoing TMS study. Subjects in that study 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 5 out of 10 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, subjects were instructed to use the computer mouse to trace an outline of the perceived phosphene onto the black computer screen with a grey central fixation cross using MATLAB PsychToolbox code. Following the drawing of at least 10 outlines, each outline was fit to an ellipse using *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. Therefore, stimuli locations are individually determined and unique for each subject in that rTMS study. For the purposes of this behavioral-only control experiment., subjects were randomly matched to the stimuli locations determined for subjects who completed the phosphene localization task and the rTMS version of the experiment.

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

Task Procedure

The flow of the experimental task is outlined in Figure 1. 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. 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 (Median Degree of Visual Angle = 10.03; SD = 4.13). 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. Subjects 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 0.3s. A green cross indicated that the response was within 15° of the target, a yellow cross indicated that the response was between 15° and 30° of the target, and a red cross indicated that the response was greater than 30° away from the target. 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 subjects 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 subjects completed 2-3 blocks in each session.

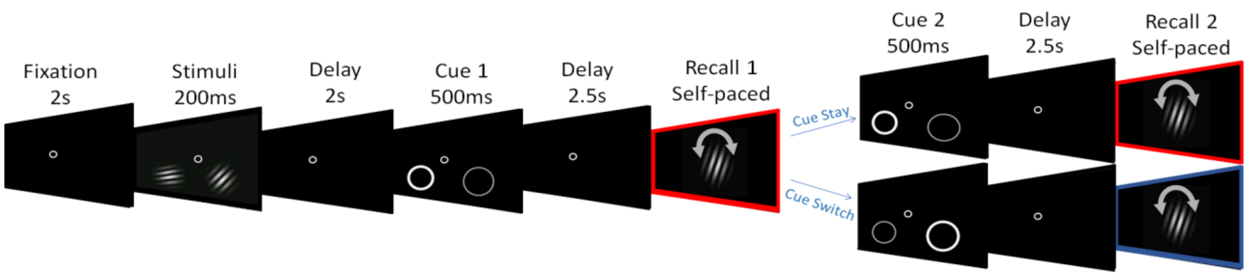


Figure 1: Illustrated task flow diagram. Gabor stimuli were briefly presented in the lower right and lower left visual hemifield. Following a delay, a bold white outline at the stimulus location served as the 100% valid retrocue. A random Gabor patch was presented at central fixation after another delay, and subjects were tasked with rotating the patch to match the cued orientation. Once their response was submitted, the task continued in one of two directions. Either the same stimulus was cued again for a second recall test (this is called a “stay” trial), or the originally uncued stimulus could be cued for the second recall test (a “switch” trial).

Data Quality Checks

To compare the precision of memory (measured as the standard deviation of errors) across the conditions, errors from each subject were calculated as the difference between the target orientation and the response orientation (Figure 2). For each recall condition per subject, the errors were converted to z-scores using the *zscore* function in MATLAB, and any z-score greater than 3 or less than -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 analysis enabled us to get a more accurate measure of memory precision. A total of 1.3% responses were 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 subject. The *boxplot* function in R Studio was then used across all subjects to determine any outliers in the dataset; two subjects were determined to be outliers and removed from the error analysis comparing behavioral performance across recall conditions. Thus, data from 33 subjects were used in the behavioral analyses.

For the mixture model analyses, all trials for the remaining subjects were included because the models attempt to separate errors by different parameters so are able to account for outliers. One subject 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 subject were not included in the group level analysis, leaving data from 32 subjects to be included in the mixture model analyses.[4]

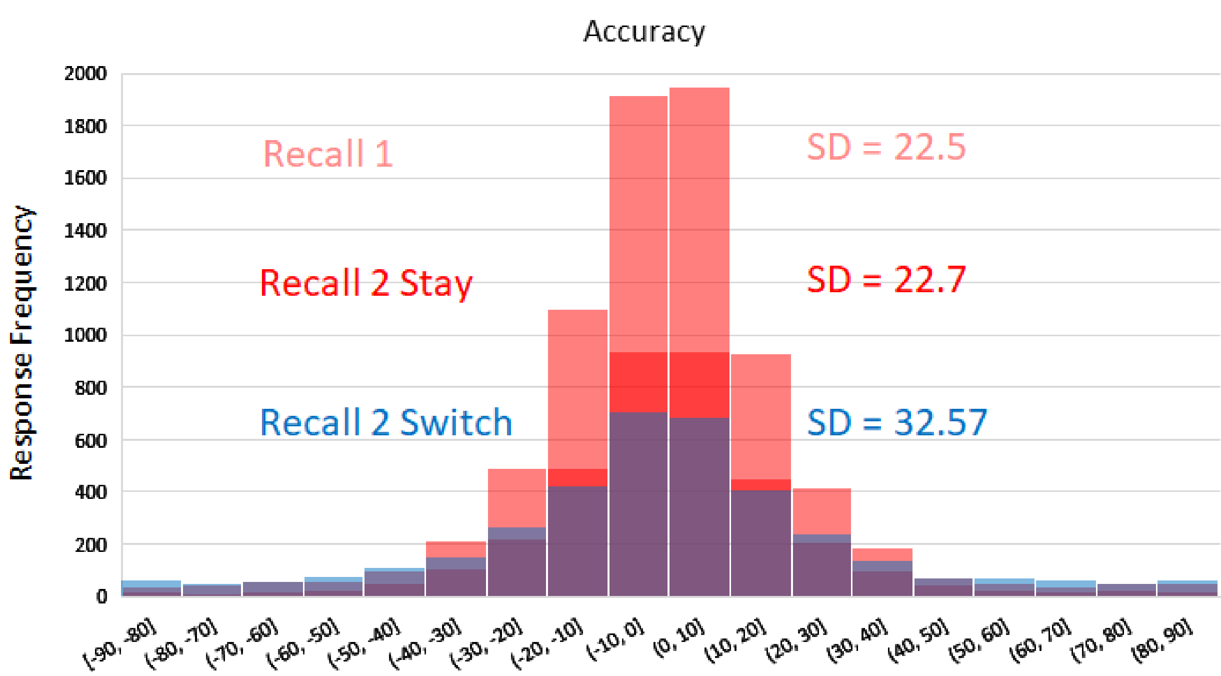


Figure 2. The frequency of recall errors and the standard deviation (memory precision) in degrees relative to the target orientation for each condition (recall 1, recall 2 stay, and recall 2 switch) for all trials and all subjects. Memory precision was similar for recall 1 and recall 2 stay trials, while 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.

[4] 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, rs = 0.05, 0.04, and 0.05, respectively, ps > 0.25.

Data Analysis

Errors were calculated by taking the difference between the target orientation and the response orientation. 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 orientation may have had on the response. To ensure a normal distribution of recall errors for each condition, a log10 transformation was used, and Shapiro-Wilks tests confirmed normality (ps > 0.1442). Analysis was conducted in R Studio.

The calculated, non-transformed error scores were used to perform mixture model analyses in MATLAB (R2016b, MathWorks, Natick, MA) using the MemToolbox (Suchow, Brady, Fougnie, & Alvarez, 2013; memtoolbox.org; Bays, Catalao, & Husain, 2009; Bays & Taylor, 2018; Peters et al., 2019; Zhang & Luck, 2008)). The models served as a means 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. Two different models were utilized: the Standard Mixture Model and the Swap Model. The Standard Mixture Model, based on Zhang and Luck (2008), used the distance of a response from the target value to determine the probability that the error of that response reflects either the precision (reflected by the standard deviation, SD) of the subject's memory for the target item or 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:

$$\rho(\hat{\theta}) = (1-\gamma)\phi_{\delta}(\hat{\theta}-\theta) + \frac{\gamma}{2\pi}. \quad (1)$$

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, standard deviation = σ, Bays et al., 2009).

The Swap Model 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. In other words, the swap error rate indicates the probability that a participant submitted a response that reflects a memory of the uncued item rather than the cued item by taking into account the accuracy of any given response relative to the nontarget item. The Swap Model is described by the equation:

$$\rho(\hat{\theta}) = (1 - \gamma - \beta)\phi_\delta(\hat{\theta} - \theta) + \frac{\gamma}{2\pi} + \left(\frac{\beta}{m}\right)\sum_i^m \phi_\delta(\hat{\theta} - \theta_i^*) \qquad (2)$$

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 subject. This allowed investigation of how memory changed when items were switched from an unprioritized state to a prioritized state within subjects. The fits of each model were compared using the Akaike Information Criterion (AIC). R Studio was used to perform all of the statistical tests on the model parameters and comparisons. T-tests comparing the three conditions were corrected for multiple comparisons with the Bonferroni correction resulting in a critical value of $p = 0.0167$).

We conducted analyses to test the influence of the nontarget distractor item on target responses across the three recall conditions. We included all trials in this analysis but removed the data of the two subjects previously deemed outliers for the average accuracy analysis since the same independent variable was used. The error for a trial was given a negative or positive value based on whether the response was between the target and nontarget item (and thus the error was committed toward the nontarget, 'attraction') or whether the target was between the response and the nontarget (meaning that the error was committed away from the nontarget, 'repulsion') respectively. Shapiro-Wilkes tests confirmed that the response errors for each recall condition were normally distributed ($p = 0.70$, $p = 0.35$, and $p = 0.28$, respectively). Finally, two ANOVAs were run in R studio to determine the relationship among response bias, recall condition, and distance between the target and nontarget line orientations between recall 1 and recall 2 stay trials and between recall 1 and recall 2 switch trials.

Results

To determine the consequences of holding information in an unprioritized state, we compared the average absolute value of the recall error across the three recall conditions (recall 1, recall 2 stay, and recall 2 switch). Recall error was higher on recall 2 switch trials (mean = 22.7, SD = 8.06) than on both recall 1 trials (mean = 14.2, SD = 5.19; $t(32) = -14.68$, $p < .001$) and recall 2 stay trials (mean = 14.6, SD = 5.77; $t(32) = -13.50$, $p < .001$), but there was no difference between recall 1 trials and recall 2 stay trials ($t(32) = -0.69$, $p = 0.50$) (Supplemental Figure 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 ran mixture model analyses (see Methods) on the data in order to better understand the source(s) of the performance 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 subjects (as in Bays & Taylor, 2018). As a whole, the Swap Model was preferred over the Standard Mixture Model for all three recall conditions: recall 1 (Mean Difference = 15.33, $p < .001$); recall 2 stay (Mean Difference = 6.77, $p < .01$); and recall 2 switch (Mean Difference = 9.45, $p < .01$).

Parameter Differences

The main purpose of this study was to elucidate the consequences of shifting visual WM items out of and back into the focus of attention, and if the no-longer relevant item biased recall of the target item. To determine what effects, if any, the shifting of attention has on memory, 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 paired t-tests were performed for each parameter to compare all three recall conditions.

Precision. As predicted, there was a statistically significant decline in the precision parameter between recall 2 switch trials and 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 ($t(31) = -0.70$, $p = 0.49$, Figure 3).

Guess Rate. The guess rate was higher for recall 2 switch trials than 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 difference between recall 1 and recall 2 stay trials ($t(31) = -1.94$, $p = 0.06$, Figure 3).

Swap Error Rate. As predicted, the swap error rate was higher for recall 2 switch trials than both recall 1 trials (one-tailed, $t(31) = -6.25$, $p < .01$) and recall 2 stay trials (one-tailed, t(31) = -4.83, p < .017), and there was no difference in swap error rate between recall 1 and recall 2 stay trials (two-tailed, t(31) = -2.18, p = 0.04, Figure 3).

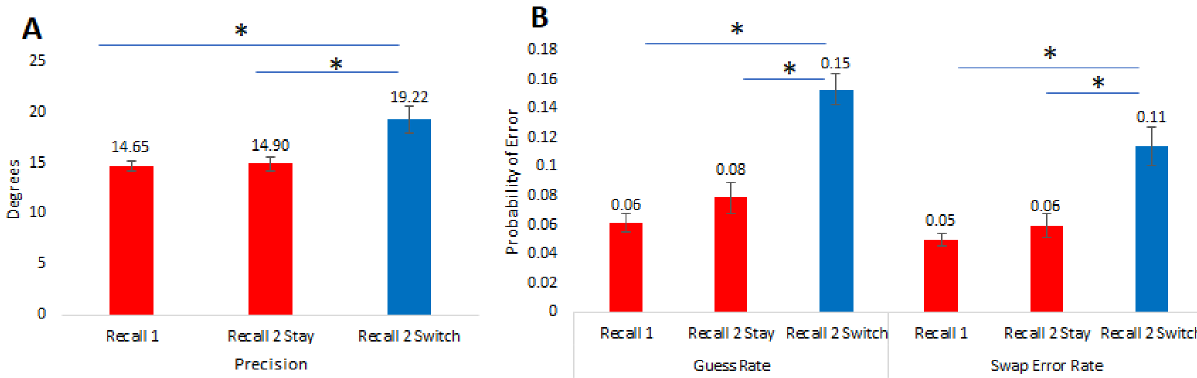


Figure 3: Average Swap Model parameter values. (a) The average precision parameter for the three recall conditions with lower values indicating better precision. Memory was less precise for recall 2 switch trials compared to both recall 1 and recall 2 stay trials. (b) The average guess rate and average swap error rate parameters for the three recall conditions. 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 and * reflects a statistically significant difference with p < 0.001.

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.

Nontarget Bias Analysis: Repulsion Effects

To elucidate the source of the differences between the recall conditions we investigated the role 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. Recall 2 switch trials exhibited greater repulsion than both recall 1 and recall 2 stay trials (t(32) = 3.52, p < 0.005) and recall 2 stay (t(32) = 3.00, p < 0.01, respectively), while no difference in bias existed between recall 1 and recall 2 stay trials (t(32) = 0.15, p = 0.88) (Supplemental Figure 2).

As in Bae and Luck (2017), we calculated the amount of bias as a function of the difference between the target and nontarget orientations. Trials were binned around three orientation differences centered at 22.5, 45, and 67.5 degrees of difference between the target and nontarget stimuli.[5] We compared the bias in response error for the three bins for the three recall conditions with ANOVAs and Bonferroni-corrected paired t-tests. There were significant effects of bin for each recall condition ($F\text{s}(2,64) > 4.88$, $p\text{s} < 0.05$), but the amount of bias was larger on recall 2 switch trials than both recall 1 and recall 2 stay trials ($F\text{s}(2,64) = 8.22$, $p\text{s} < 0.001$), and there was no difference between recall 1 and recall 2 stay trials (Figure 4). The amount of bias between recall conditions was not significant when there were large (~67.5 degrees) differences between the target and non-target ($p\text{s} > .45$), but it was significant for smaller differences ($p\text{s} = .0026$ and .0053 for the 45 degree bin for recall 2 switch vs. recall 1 and recall 2 stay, respectively, and .0177 and .0179 for the 22.5 degree bin, though the difference failed to survive Bonferroni correction of .0167 for the latter bin).

[5] Bae and Luck (2017) presented unidirectional line orientations with the potential for 180 degrees of difference between the two stimuli. However, in this experiment we presented bidirectional line orientations, so the greatest possible difference between target and nontarget was 90 degrees. Therefore, our replication was limited to trials in which the difference between target and nontarget was less than or equal to 90 degrees.

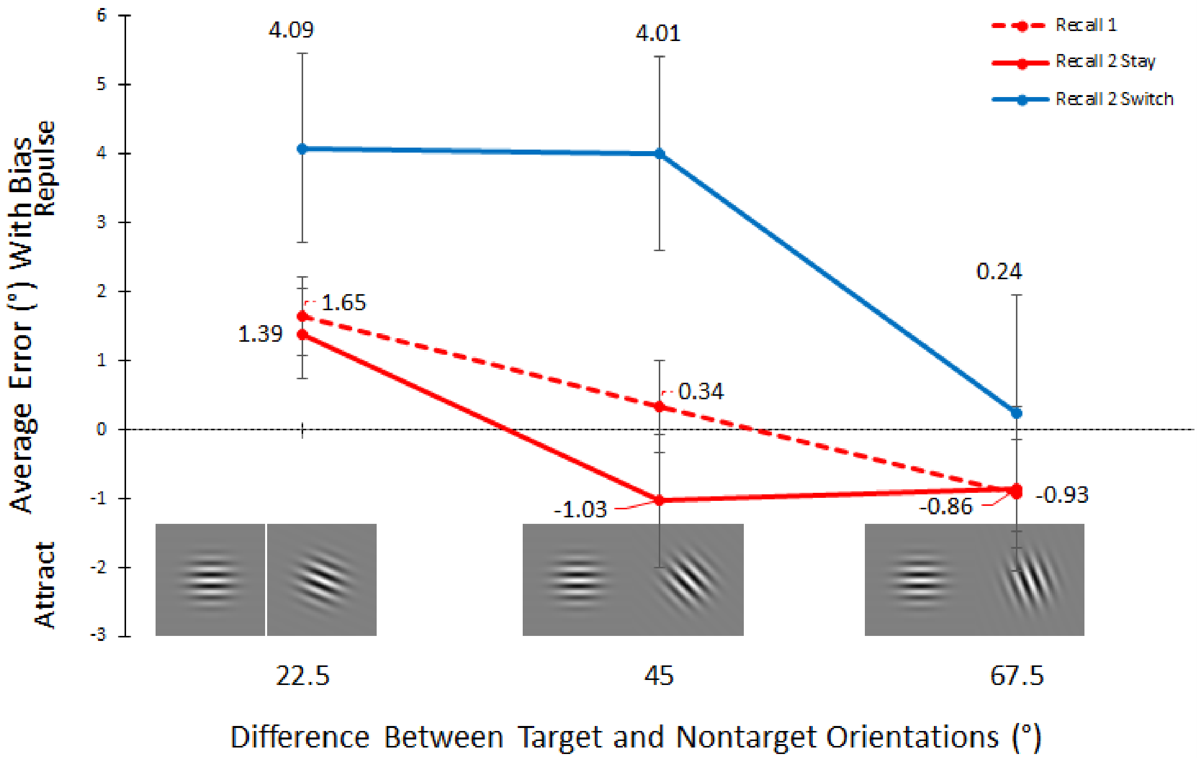


Figure 4: Average response error bias as a function of distance between orientation stimuli. Response errors were calculated as the degree of difference from the target orientation towards (negative) or away from (positive) the nontarget item, and the average response bias across participants was plotted for each recall condition. Bins were created using trials in which the difference between the stimuli were ± 7 degrees from 23, 45, and 68 degrees, respectively. There was no difference in bias between recall 1 and recall 2 stay trials, but there was a significant difference in bias between recall 1 and recall 2 stay trials. There was also a significant difference in bias across the bins for all three recall conditions. Error bars reflect 1 standard error of the mean.

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, 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 items in WM that are no-longer-relevant for ongoing behavior are actively deleted from WM, these no-longer-relevant 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 expected pattern based on the active-deletion hypothesis. Following the second retrocue, the uncued item was no longer relevant and, therefore, should have been deleted from WM and resulted in *less* bias for responses on recall 2 stay and recall 2 switch trials. The present results, which replicate and extend those reported by Bae and Luck (2017), suggest that no-longer-relevant items were not deleted from WM following the second retrocue.

The results converge with those of Bae and Luck (2017) despite important differences between the experimental paradigms. Stimuli were presented sequentially at central fixation in that study whereas stimuli were presented simultaneously in the lower left and right hemifields in the present study. These are not trivial methodological differences. There was considerably more overlap in the cortical areas that processed the visual stimuli in Bae and Luck's 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 (Scotti et al., 2021). 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 degrees clockwise/counterclockwise from the first stimulus). Also, 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. 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 Bae & Luck's paradigm). Nevertheless, the data from both studies showed strikingly similar evidence that items that should have been deleted from WM can bias retrieval of a target item in WM, directly contradicting the active-deletion hypothesis, and diverging from evidence supporting the notion that no-longer-relevant items do not influence retrieval of target items in WM (i.e., Rose et al., 2016; Fulvio & Postle, 2020).

One possible explanation for this pattern of results is that, when trying to remember two line orientations, participants may have encoded the two distinct orientations as an 'ensemble representation'. As Bae and Luck suggested, participants could bind the two distinct orientation objects into one ensemble representation, with both orientations bound together as an angle or clock hands, for example. Anecdotal evidence from post-experimental debriefing of our participants supports this interpretation. Although the two gabor orientations were presented separately in the lower left and right hemifields, most participants reported encoding the two as an angle by projecting the lines out to their intersecting point. Encoding the two objects as a bound angle changes the nature by which the no-longer-relevant item can be deleted and may prevent the uncued item from dropping from focal attention compared to paradigms with more distinct (e.g., categorical) stimuli, such as a face paired with a word or direction of motion, as in Rose et al. (2016) (see also, Fulvio & Postle, 2020). 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 the second retrocue.

The differences between the categorical vs. bound nature of the representations can explain the differences between the results reported here (and in Bae & Luck, 2017) and those from Rose et al. (2016) and Fulvio & Postle (2020). Those studies 1) required participants to encode items from distinct categories that were spatially separated during encoding, and 2) tested WM with recognition probes from the same category as the cued item. The stimuli were also trial-unique, meaning that they were never repeated, so the match/nonmatch recognition decisions could likely be made based on the strength of a familiarity signal resulting from a template-matching process (Clark & Gronlund, 1996; Wickelgren & Norman, 1966; Yonelinas, 1994). In contrast, the present experiment required participants to encode two low-level visual features (line orientations) that could be bound into a unitary object.

Nevertheless, clarifying the exact source of the differences in results is an important direction for future research examining the dynamics of WM. A limitation of this study is that no neuroimaging or neurostimulation methods were used to indirectly "observe" the activation status of items held in WM (Lee & Baker, 2016; 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 bound as ensembles. Our approach in analyzing bias from no longer relevant items can serve as a guide for studies attempting to elucidate the nature of WM representations and how they are influenced by other items in memory.

Acknowledgments

This manuscript 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 authors would like to thank Jori Waner, Isaiah Metcalf, Cheick Diallo, Yuki Yoshioka, Lauren Crowe, and Bernadette Widjaja for contributing to data collection and analysis. The authors declare no conflicts of interest. The Data are publicly available at CurateND: https://doi.org/10.7274/r0-mq7c-7m28.

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, and Psychophysics*, *79*(8), 2376–2395.

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).

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

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

Clark, S. E., & Gronlund, S. D. (1996). Global matching models of recognition memory: How the models match the data. Psychonomic bulletin & review, 3(1), 37-60.

Conway, A. R. A., Jarrold, C., Kane, M. J., Miyake, A., & Towse, J. N. (2012). Variation in Working Memory. In *Variation in Working Memory*. Oxford University Press.

Cowan, N. (1998). Attention and Memory: An Integrated Framework. In *Attention and Memory: An Integrated Framework*. Oxford University Press.

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.

Duarte, A., Hearons, P., Jiang, Y., Delvin, M. C., Newsome, R. N., & Verhaeghen, P. (2013). Retrospective attention enhances visual working memory in the young but not the old: An ERP study. *Psychophysiology*, *50*(5), 465–476.

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, and Psychophysics*, *77*(8), 2562–2569.

Gunseli, E., van Moorselaar, D., Meeter, M., & Olivers, C. N. L. (2015). The reliability of retro-cues determines the fate of noncued visual working memory representations. *Psychonomic Bulletin and Review*, *22*(5), 1334–1341.

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

Heuer, A., & Schubö, A. (2016). Feature-based and spatial attentional selection in visual working memory. *Memory and Cognition*, *44*(4), 621–632.

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.

Lee, S. H., & Baker, C. I. (2016). Multi-voxel decoding and the topography of maintained information during visual working memory. Frontiers in systems neuroscience, 10, 2.

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.

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

Makovski, T., Sussman, R., & Jiang, Y. V. (2008). Orienting Attention in Visual Working Memory Reduces Interference From Memory Probes. *Journal of Experimental Psychology: Learning Memory and Cognition*, *34*(2), 369–380.

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. (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.

Oberauer, K., Lewandowsky, S., Awh, E., Brown, G. D. A., Conway, A., Cowan, N., Donkin, C., Farrell, S., Hitch, G. J., Hurlstone, M. J., Ma, W. J., Morey, C. C., Nee, D. E., Schweppe, J., Vergauwe, E., & Ward, G. (2018). Benchmarks for models of short-term and working memory. *Psychological Bulletin*, *144*(9), 885–958.

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).

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.

Samaha, J., Sprague, T. C., & Postle, B. R. (2016). Decoding and reconstructing the focus of spatial attention from the topography of alpha-band oscillations. *Journal of Cognitive Neuroscience*, *28*(8), 1090–1097.

Scotti, P. S., Hong, Y., Leber, A. B., & Golomb, J. D. (2021). Visual working memory items drift apart due to active, not passive, maintenance. *Journal of Experimental Psychology: General*.

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.

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, 1–8.

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.

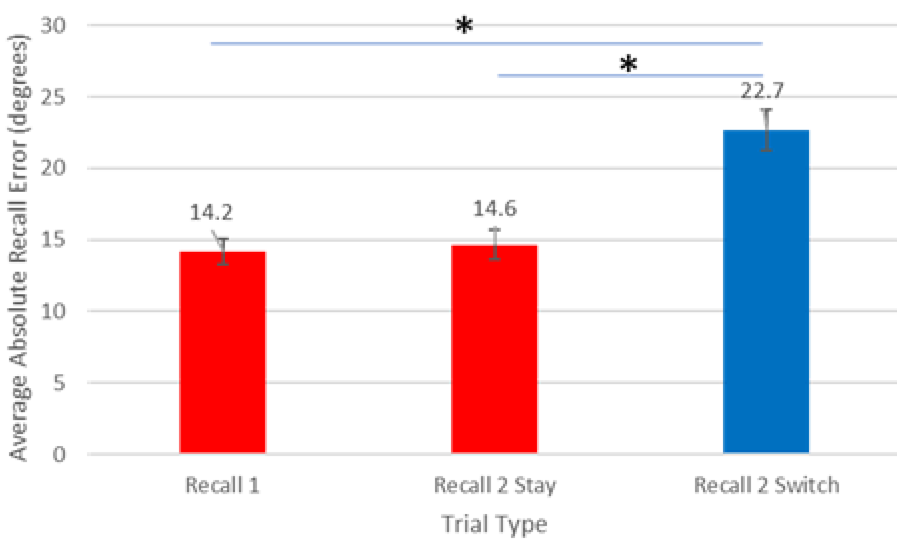
Wickelgren, W. A., & Norman, D. A. (1966). Strength models and serial position in short-term recognition memory. Journal of Mathematical Psychology, 3(2), 316-347.

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.

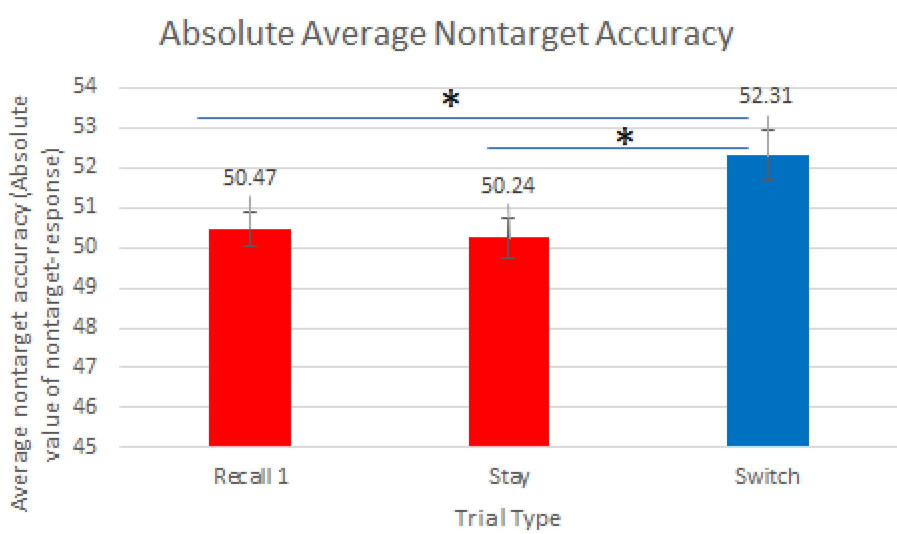
Yonelinas, A. P. (1994). Receiver-operating characteristics in recognition memory: evidence for a dual-process model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(6), 1341.

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

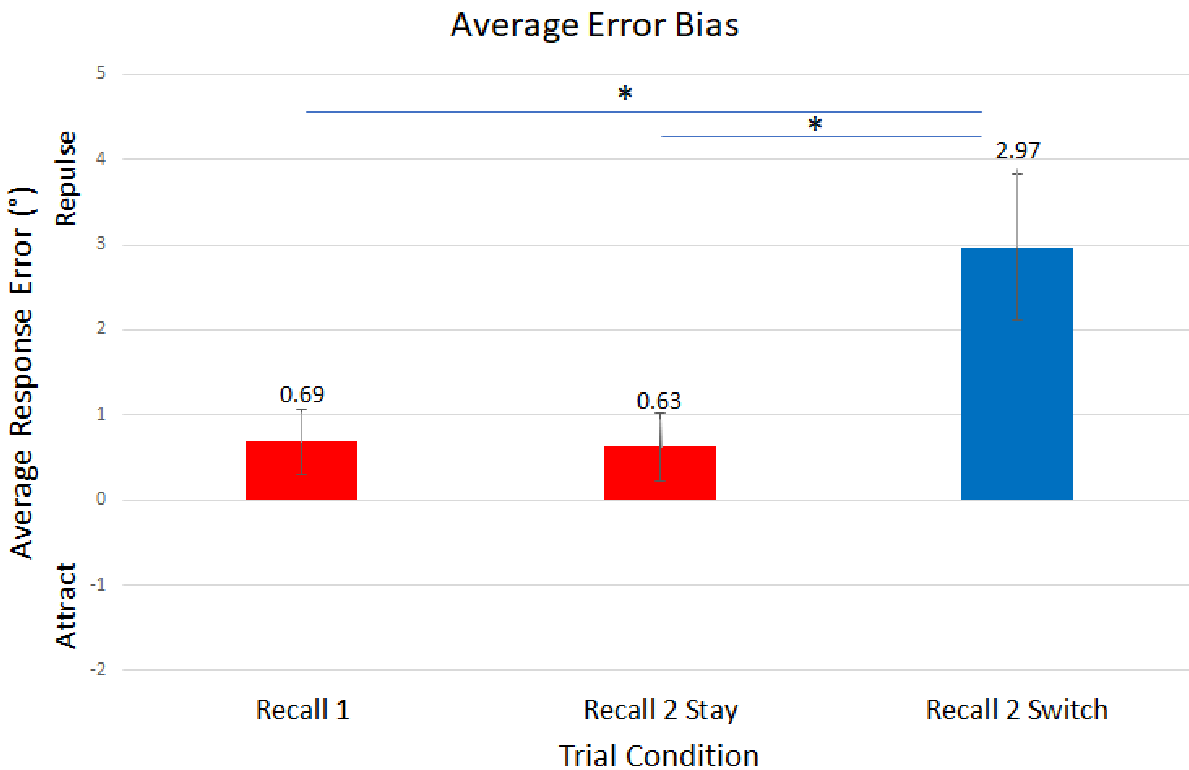
Supplemental Materials



Supplemental Figure 1: Average absolute value of recall error by recall condition. Recall error was calculated as the difference between the response orientation and the target orientation. Recall error was significantly worse for recall 2 switch trials compared to both recall 1 trials and recall 2 stay trials, indicating a loss in memory fidelity for items shifted to the unattended state. Error bars reflect 1 standard error of the mean, and * reflects a statistically significant difference with $p < 0.001$.



Supplemental Figure 2: Average absolute value of recall error relative to the nontarget by recall condition. The absolute value of the distance between the response and the uncued, nontarget item was calculated and averaged across all subjects for responses in each recall condition. Responses on recall 2 switch trials were further away from the nontarget compared to both recall 1 and recall 2 stay trials (both ps < 0.01). This trend suggests that some factor or factors influence responses on recall 2 switch trials more than on the other two trial conditions. * reflects statistically significant differences.



Supplemental Figure 3: 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. While all trial conditions exhibited a net repulsive bias away from the nontarget item, this repulsion was greater on recall 2 switch trials than both recall 1 and recall 2 stay trials (both ps < 0.001). Error bars reflect 1 SEM and * indicate statistically significant differences.