Journal of Experimental Psychology: General

Visual Working Memory Items Drift Apart Due to Active, Not Passive, Maintenance

Paul S. Scotti, Yoolim Hong, Andrew B. Leber, and Julie D. Golomb

Online First Publication, May 20, 2021. http://dx.doi.org/10.1037/xge0000890

AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: General

ISSN: 0096-3445

https://doi.org/10.1037/xge0000890

Visual Working Memory Items Drift Apart Due to Active, Not Passive, Maintenance

Paul S. Scotti, Yoolim Hong, Andrew B. Leber, and Julie D. Golomb
Department of Psychology, The Ohio State University

How are humans capable of maintaining detailed representations of visual items in memory? When required to make fine discriminations, we sometimes implicitly differentiate memory representations away from each other to reduce interitem confusion. However, this separation of representations can inadvertently lead memories to be recalled as biased away from other memory items, a phenomenon termed repulsion bias. Using a nonretinotopically specific working memory paradigm, we found stronger repulsion bias with longer working memory delays, but only when items were actively maintained. These results suggest that (a) repulsion bias can reflect a mnemonic phenomenon, distinct from perceptually driven observations of repulsion bias; and (b) mnemonic repulsion bias is ongoing during maintenance and dependent on attention to internally maintained memory items. These results support theories of working memory where items are represented interdependently and further reveals contexts where stronger attention to working memory items during maintenance increases repulsion bias between them.

Keywords: working memory, mnemonic bias, repulsion bias, attentional competition, hierarchical bayesian mixture model

Supplemental materials: https://doi.org/10.1037/xge0000890.supp

Task-irrelevant information such as the visual similarity between memory items (e.g., Golomb, 2015), ensemble statistics (e.g., Brady & Alvarez, 2011), and spatial context (e.g., Awh & Jonides, 2001; Jiang et al., 2000) can all bias how a memory item is remembered mere seconds after encoding. It has been argued that such biases emerge due to the memory system optimally combining various sources of information to increase overall performance (e.g., Carpenter & Schacter, 2017; Guerin et al., 2012; Huttenlocher et al., 2000; Newman & Lindsay, 2009; Schacter et al., 2011; Yoo et al., 2018). For instance, in a task where you memorize several squares and then report the size of a target square that is cued after a blank delay, it might be beneficial for the memory system to calculate the average size during encoding. This average size can act as a reference point that biases your report toward the average feature when uncertain (Brady & Alvarez, 2011). While this would induce systematic biases in memory reports throughout the experiment, it could also benefit overall performance in terms of absolute error away from the target feature.

Interestingly, such memory biases can be adaptive both toward and away from a reference point. The previous example described memory reports biased toward a reference point (i.e., attraction bias), a phenomenon thought to arise from the combination of item-level and group-level information, perhaps arising from a Bayesian process of combining prior information with an uncertain stimulus (e.g., Brady & Alvarez, 2011; Brady et al., 2018; Hemmer & Steyvers, 2009; Huttenlocher et al., 2000). A bias away from a reference point (i.e., repulsion bias) can also be observed. For both attraction and repulsion bias, any task-relevant feature can serve as a reference point, encompassing properties such as the feature of a competing memory item (e.g., Golomb, 2015), ensemble statistics (Brady & Tenenbaum, 2013; de Fockert & Wolfenstein, 2009; Haberman & Whitney, 2009), statistical regularities (Honig et al., 2020), subjective category labels (Bae et al., 2015; Huttenlocher et al., 1991), and the perceptual average of sequential stimuli (Bae & Luck, 2017; Huang & Sekuler, 2010).

Repulsion bias could serve to help minimize confusability with a reference point by subtly biasing the representation in feature space away from the reference point, and is most often observed in tasks that require a few easily confusable items to be maintained in fine detail (Bae & Luck, 2017; Chen et al., 2019; Golomb, 2015). For example, if one colored square was light blue and

Paul S. Scotti https://orcid.org/0000-0003-4912-8809
Julie D. Golomb https://orcid.org/0000-0003-3489-0702

Paul S. Scotti, Yoolim Hong, Andrew B. Leber, and Julie D. Golomb contributed to theoretical motivation, study design, analysis plan, and writing of the article. Paul S. Scotti drafted the article, programmed the experiments, and performed data collection/analyses.

The authors declare no competing interests.

This study was funded by grants from the National Science Foundation (NSF DGE-1343012 to Paul S. Scotti; NSF BCS-1632296 to Andrew B. Leber; NSF BCS-1848939 to Julie D. Golomb and Andrew B. Leber) and the National Institutes of Health (R01-EY025648 to Julie D. Golomb). Analyses reported in this publication were supported by an allocation of computing resources from the Ohio Supercomputer Center (1987).

Correspondence concerning this article should be addressed to Paul S. Scotti, Department of Psychology, The Ohio State University, 1835 Neil Avenue, Columbus, OH 43210, United States. Email: scottibrain@gmail.com

another was dark blue, it may be optimal for the memory system to push apart these representations in color space to ensure that these squares are not confused with each other. Repulsion bias could be implemented in the brain by the neural mechanisms of lateral inhibition (Johnson et al., 2014; Johnson et al., 2009; Wei et al., 2012) and/or optimal gain (Navalpakkam & Itti, 2007; Scolari & Serences, 2009). We discuss these theories more in the General Discussion.

Attraction and repulsion bias can be observed in the absence of any mnemonic processing. For example, in the direction illusion, subjects estimate the direction of two transparent and overlaid random dot patterns and mutual attraction or repulsion is observed depending on the relative difference in direction (e.g., Blakemore et al., 1970; Levinson & Sekuler, 1976; Marshak & Sekuler, 1979; Mather, 1980; Rauber & Treue, 1998; Wiese & Wenderoth, 2007; Yo & Wilson, 1992). Likewise, the tilt illusion can perceptually induce repulsion or attraction bias: Subjects view an oriented grating that is surrounded by another oriented grating, and the center grating is perceived to be either repulsed by or attracted to the surrounding grating depending on the relative difference in orientation (e.g., Gibson & Radner, 1937; O'Toole & Wenderoth, 1977). Direction aftereffects (e.g., Hiris & Blake, 1996; Wenderoth & Wiese, 2008; Wiese & Wenderoth, 2007) and tilt aftereffects (e.g., Gibson, 1937; Wenderoth & Johnstone, 1988) are extensions of these illusions, where visual adaptation induces retinotopically specific perceptual biases on subsequent visual input.

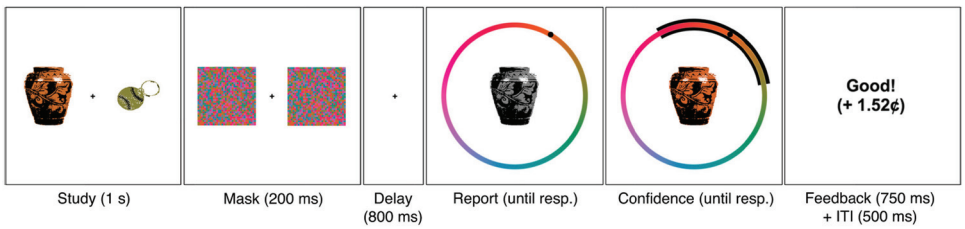
These perceptual illusions might be related to the attraction/repulsion biases observed in previous working memory studies. Previous working memory studies have used paradigms where items are encoded and tested in the same spatial location, such that working memory maintenance might have relied on retinotopically specific, sustained sensory activation (Czoschke et al., 2020). In other words, persistent neural firing in sensory areas after stimulus offset could induce visual adaptation effects similar to those observed in the aforementioned perceptual illusions.

In the current study, we first tested whether attraction/repulsion bias can be observed mnemonically, in the absence of a purely perceptual explanation, by testing for such biases in a nonretinotopically specific working memory paradigm. Specifically, we designed an experiment where memory items were encoded and tested in different spatial positions. Instead of using simple geometric shapes that are cued based on spatial location (e.g., Golomb, 2015) or temporal position (e.g., Bae & Luck, 2017), we used real-world objects that were cued based on object identity (see Figure 1). Participants viewed two real-world objects with colors sampled either 45 ° or 90 ° apart in color space, and then reported the color of the cued object after a short working memory delay.

Using this paradigm, we conducted a series of four preregistered experiments aiming to provide a better understanding of how and when working memory representations interact with each other. Working memory is capacity-limited (Luck & Vogel, 2013), and the canonical theory of working memory capacity is that all items are represented independently (Luck & Vogel, 1997; Zhang & Luck, 2008). Mnemonic bias provides support for interdependent memory items, in line with more recent theories of working memory (e.g., Brady & Alvarez, 2015; Johnson et al., 2014; Oberauer & Lin, 2017). We tested if mnemonic bias is present independent of perceptual bias, and if so, when it emerges, by manipulating the duration and active versus passive nature of the working memory delay. If representations do interact with each other in working memory, an important question is when biased representations emerge during the stages of encoding, maintenance, and retrieval. We offer three (nonexclusive) possibilities. (a) The target memory is biased as soon as or very soon after the study array disappeared (i.e., bias during encoding). (b) The target memory becomes biased during the working memory delay (i.e., bias during maintenance). (c) The target memory becomes biased after the participant is told which memory item needs to be reported (i.e., bias during retrieval).

In Experiments 1–3, all trials had identical encoding and retrieval demands, but varied in the duration of the working memory delay. To preview our primary results, we did not observe a credible main effect of repulsion bias with a blank working memory delay of 1 s, but we did observe credible repulsion bias with a blank working memory delay that was 3 s long, suggesting that mnemonic bias can be observed independent of perceptual bias, and it emerges during the maintenance period. In Experiment 4, we attempted to disambiguate two potential reasons why mnemonic bias might emerge over longer delays, asking whether active attention to the items in memory during the delay is required. Previous studies have found that working memory

Figure 1
Example Trial Sequence for Experiment 1



Note. Participants were instructed to memorize the colors of both real-world objects. Following a blank interval, participants recreated the original color of one of the two objects (randomly selected), cued by presenting the object in the center of the screen in grayscale. Objects were displayed in grayscale until mouse movement, at which point the objects color dynamically adjusted to match the color closest to the mouse pointer. After clicking to confirm their best guess, participants highlighted the smallest range of colors that they believed contained the original color. Every test trial ended with general feedback and bonus information. See the online article for the color version of this figure.

representations deteriorate over time (Barrouillet & Camos, 2009; Barrouillet et al., 2012; Pertzov et al., 2013; Vergauwe et al., 2009), and poorer quality representations may lead to interitem confusability, prompting the memory system to distinguish the representations in feature space (Bae & Luck, 2017; Chunharas et al., 2019). Thus, one might expect repulsion bias to increase with longer memory delays, even—or perhaps more so—when people are not actively attending to the items during the delay. However, an alternative account predicts that active maintenance of competing representations induces repulsion bias over time, and stronger repulsion bias should be observed during longer delays only when the items are actively attended during the delay. In Experiment 4, we compared a 3-s blank working memory delay (where the task was to actively maintain the working memory items) to a 3-s delay that involved a separate filler task (attend to something else during the memory delay) to disambiguate these accounts, ultimately supporting the active maintenance account, allowing us to better inform theories of working memory.

Experiment 1: 1-S Maintenance Duration

We first tested for memory distortions in a visual working memory experiment where memory items were encoded and tested in different spatial positions. We started with a 1-s blank working memory delay in Experiment 1, to facilitate comparison with more perceptually based working memory studies reporting shift errors (also known as feature bias, either attraction or repulsion from a nontarget feature) and swap errors (reporting a nontarget feature; e.g., Golomb et al., 2014).

Regarding shift errors, if repulsion and/or attraction bias is observed, this would provide initial evidence that such distortions are not retinotopically specific and can be attributed to a mnemonic, not a perceptual, explanation. We were also curious whether repulsion might be stronger when objects are separated in color space by 45 °, compared with 90 °. The previously discussed perceptual illusions and relational representation model claim that the relative difference in feature space is an important factor that can influence whether repulsion or attraction bias is observed (e.g., Bae & Luck, 2017; O'Toole & Wenderoth, 1977; Wiese & Wenderoth, 2007). We hypothesized that, if mnemonic distortions were observed, repulsion bias would be present for the 45 ° color difference and that either repulsion bias or attraction bias might be present for the 90 ° color difference.

We were also curious whether swap errors (reporting the feature of the nontarget object, also known as "misbindings" or "misassociations"; e.g., Bays et al., 2009; Bays et al., 2011) might be observed. It is possible that swap errors previously observed in visual working memory experiments (e.g., Chen et al., 2019; Dowd & Golomb, 2019; Golomb et al., 2014) were only observed because they used perceptually similar stimuli. For example, if a participant needs to memorize the colors of two squares, this is an example where the target and the nontarget items are identical except for the task-relevant color. Swap errors may be less likely to occur if the stimuli are two real-world objects, where the objects contain different task-irrelevant features (e.g., low-level features such as shape and high-level features such as semantic identity).

Method

Open Practices Statement

The rationale, method, and parts of the analyses for this and subsequent experiments were preregistered at the Open Science Framework (OSF; https://osf.io/usrxq/?view_only=6d1f0755173f43b3b6d8e8d161dd7fdc).

All analysis code and data are also available on OSF. Any analyses not mentioned in the preregistrations are declared as exploratory. Analyses reported in the main text deviated from the preregistration in the following ways: (a) We preregistered a specific nonhierarchical mixture model and stated that we may instead (or in addition) use a hierarchical Bayesian mixture model (HBMM). Because hierarchical Bayesian modeling offers substantial advantages over nonhierarchical modeling (Estes, 1956; Heathcote et al., 2000; Oberauer et al., 2017), we present the HBMM results as our primary focus in the main text, and the non-HBMM model results in the online supplemental material. In making the HBMM, we also deviated partially in the model parametrization (explained in the online supplemental material). The two modeling approaches were highly consistent in their findings. (b) For Experiments 1 and 2, when analyzing shift errors separately for the two color-difference conditions, we made an additional simplification to the within-subject HBMM model to make the model fits more reliable for this lower-powered analysis. The unsimplified model still showed consistent results, as reported in the online supplemental material in Tables S7 and S8. (c) Analyses involving confidence reports are presented in the online supplemental material, and our preregistered confidence analyses within experiments were replaced with a more powerful set of analyses across experiments (explained further in the online supplemental material).

Participants

Experiment 1 included a preregistered sample size of 50 participants (32 male, 17 female, one nonbinary; M = 37.32 years, SD = 10.51). All participants were recruited through Amazon Mechanical Turk (MTurk) and were paid $6 USD per hour (plus bonus based on performance) for the experiment, which lasted roughly 1 hr long. All participants lived in the United States, held an MTurk approval rating of ≥98%, and successfully completed over 750 MTurk tasks prior to this experiment. All participants reported normal or corrected-to-normal vision, were naive to the purpose of the experiment, and provided informed consent in accordance with The Ohio State University institutional review board. Ten participants were excluded based on preregistered exclusion criteria (>.5 guessing proportion based on a basic, nonhierarchical mixture model composed of a target and a guessing distribution; see Bays, et al., 2009; Golomb et al., 2014; Zhang & Luck, 2008).

Stimuli and Procedure

Experiment 1 was conducted online using MTurk, meaning that monitors could vary in size, viewing distance, color calibration, and so forth. We report stimulus sizes in pixels and not degrees of visual angle because of these variable environments. Figure 1 illustrates an example trial sequence. Each of 14 blocks consisted of 20 trials and 40 unique real-world objects. On each trial, two 200 × 200 pixels objects were presented to the left and right of a central 15 × 15 pixels black fixation cross. The center of each

object was 150 pixels away from the central fixation cross. All stimuli were displayed inside of a 600 × 600 pixels white square. Objects were displayed for 1 s, followed by a 200-ms mask and then an 800-ms blank interval. The mask was sized 200 × 200 pixels and composed of 400 10 × 10 pixel squares, with each square sampling a color randomly drawn from the color wheel. Participants were instructed to memorize the colors associated with every object, knowing that they would be asked to reproduce the color one of the two objects at the end of every trial.

Object stimuli were acquired from Brady et al. (2008). Objects were posterized such that pixel values could only be white, black, or a single color of interest (one of 360 RGB color values drawn from a circle of 60° radius in CIE L*a*b* color space, centered at L = 70, a = 20, and b = 38). CIE colors were converted to gamma-corrected RGB values, color-calibrated using a Chroma Meter CS-100 on a MacBook Pro. Specifically, each image was first converted to grayscale, with luminance values ranging from 0 to 255. Pixels with a luminance between 0 and 85 were colored white and pixels with a luminance between 170 and 255 were colored black. All other pixels were assigned to a color in CIE L*a*b* space. Author PS subjectively curated the objects such that they remained recognizable, contained a reasonable number of colored pixels, were not associated with a canonical color (e.g., no firetrucks because their canonical color is red), would not provoke a strong emotional reaction, and were categorically distinct from the other objects in the folder (e.g., we would not allow two exemplars of an apple).

Following the 800-ms blank interval, one of the two studied objects was randomly selected and displayed in luminance-corrected grayscale on the center of the screen. A randomly rotated color wheel (flipped on half the trials) was presented around the object. As the mouse moved around the color wheel, the initially grayscale object dynamically changed to the color closest to the mouse pointer. Participants were tasked with selecting the original color of the object by clicking on the color wheel. Following a mouse click to confirm their selected color, participants highlighted the smallest portion of the color wheel that they believed contained the original color (see Chen et al., 2019). Highlighting involved dragging with the mouse to define the start and end points of a black, highlighted region. This confidence report was our proxy for subjective memory strength, with the assumption that a larger highlighted region indicated that a participant was less certain about their memory retrieval. Confidence analyses are detailed in the onilne supplemental material. There was no time limit to respond. Following confidence reports, general feedback was presented for 750 ms, followed by a 500-ms blank intertrial interval (ITI).

A monetary bonus was presented during feedback, dependent on the subject's performance. The bonus for each trial was calculated according to (a) degrees of error (distance from the subject's reported color to the correct color) and (b) confidence range report. For the bonus calculated based on degrees of error, cents awarded equaled $1 - x/45$ (where x is absolute degrees of error), such that more fractions of a penny were awarded for less degrees of error but nothing was awarded if $x \geq 45$. For the confidence range report, cents awarded equaled $(360 - y)/359$ (where y is the confidence range where y = 360 is a highlight of the entire color wheel), such that smaller ranges awarded more money (the minimum highlighted range was 1 °). However, if the highlighted region did not contain the true original color, then no bonus was awarded for this part. The maximum bonus that could be awarded on a trial was 2 cents and the minimum bonus was 0 cents. Subjects were informed about how their bonus was calculated before starting the experiment.

On each trial, the colors of one of the two objects was randomly sampled from the color wheel, while the other object was equally likely to be sampled ±45 or ±90 ° away from the other object in color space. Participants were uninformed of this color sampling manipulation. A practice block of five trials familiarized participants with the procedure. Stimulus presentation was facilitated by a combination of HTML, CSS, and JavaScript.

Analyses

For all experiments, memory response distributions were fit using a hierarchical Bayesian mixture model (HBMM) in JAGS (Plummer, 2003), results from a nonhierarchical model are presented in the online supplemental material. HBMMs are advantageous because they provide accurate group-level and individual-level parameter estimates within a single model (Lee & Wagenmakers, 2014). In contrast to HBMMs, more traditional approaches to fitting memory response distributions have downfalls. For instance, group-level maximum likelihood estimation does not take into account individual differences. Meanwhile, individual-level maximum likelihood estimation can lead to unreliable estimates, and frequentist statistics on such estimates disregards the variability of each individual's parameter estimates. An advantage of HBMMs is that data from all participants in the study can inform individual-level estimates, allowing for more robust parameter estimates and increased statistical power without having to average data across participants (for more information see Estes, 1956; Heathcote et al., 2000; Oberauer et al., 2017). Analyses for all experiments were supported by an allocation of computing resources from the Ohio Supercomputer Center (1987). Please see the onilne supplemental material for a more complete description of our HBMM and our OSF project (https://osf.io/usrxq/?view_only=6d1f0755173f43b3b6d8e8d161dd7fdc) for the R and JAGS code for implementing the HBMM.

Each trial's memory response was first converted into an error measurement (i.e., the difference between the reported color and the correct color of the target object, measured in radians along the color wheel; we later converted from radians to degrees when reporting results). The sign of this error measurement was determined relative to the nontarget object's color: Error measurements were aligned such that the nontarget object's color was always in the positive direction (aligned to +45° or +90° on the color wheel). In this way, we could observe a mean shift in the target distribution where responses were either toward (attraction) or away from (repulsion) the nontarget color.

Memory response distributions were then fit as a mixture of three distributions: a target distribution (expressed as Ptarget), a nontarget (swap) distribution (Pswap), and a random guessing distribution (Pguess). The target distribution was a von Mises distribution (equivalent to circular normal distribution) intended to characterize memory reports where the subject correctly reported the original color of the target (with some room for error, characterized by the concentration parameter). The nontarget, or swap, distribution, was a von Mises intended to characterize memory reports where the subject mistakenly reported the color of the nontarget item (Bays et al., 2009; Golomb et al., 2014; Scotti et al., 2021). The random guessing distribution was characterized by a

circular uniform distribution and was intended to characterize memory reports where the subject was randomly reporting a color on the color wheel.

We implemented a similar hierarchical three-component mixture model as Oberauer et al. (2017), but modified it to allow the center of the target distribution to flexibly shift up to 15 ° in either direction, such that we could assess repulsion bias (negative shift) or attraction bias (positive shift). Allowing the target distribution to flexibly shift has previously been used in nonhierarchical models to observe subtle attraction or repulsion bias (e.g., Chen et al., 2019; Golomb, 2015; Golomb et al., 2014). We restricted the target distribution to shift a maximum of ±15 ° because this would ensure that nontarget responses (45° or 90° from the target color) were not accidentally attributed to the target distribution. We also fit separate concentration parameters (equivalent to the inverse variance and often called "precision") to the target and swap distributions to allow for the possibility that swap errors may be associated with lower precision.

The HBMM can be described according to its hierarchical levels (in descending order): group-level, condition-level, and subject-level. We used a HBMM with only two levels (group-level and subject-level) when testing for overall, group-level effects (e.g., overall repulsion bias across subjects) and we used a HBMM with three levels when testing for condition-level differences (e.g., difference in repulsion magnitude between color-distance conditions). At each level, there are parameters that define the relative proportion of target responses, relative proportion of swap responses, relative proportion of random guessing, shift in mean (bias) of the target distribution, and precisions of the target and swap distributions. The parameters of the lower levels have priors that are based on the respective parameter from the immediate higher level; that is, the condition-level has parameters with priors depending on the respective group-level parameters and the subject-level has parameters with priors from the respective condition-level parameters (or group-level for the two-level model). The full details and formulas for our HBMM can be found in the online supplemental material.

For each model fit, we collected 15,000 postconvergence samples and used the posterior distributions to compute the maximum a posteriori (MAP) group-level and individual-level parameter estimates. In computing the MAP estimate we used Silverman's kernel density estimation (Silverman, 1986) to obtain the mode of the posterior distribution. (Note that the PTarget, PGuess, and PSwap parameters summed to 1 within each of the 15,000 samples, but they did not always sum to 1 in the reported MAP estimates due to this process. The nonhierarchical model results in the online supplemental material report maximum likelihood estimation and do sum to 1, which show consistent findings to the HBMM). We verified convergence with the Gelman-Rubin convergence diagnostic (Gelman & Rubin, 1992).

For each parameter, we further use the 15,000 postconvergence samples to calculate the 95% highest density interval (HDI). 95% HDIs indicate that the true parameter value has a 95% probability of lying within this interval. Values outside the intervals may be considered sufficiently implausible (Lindley, 1965). For example, when we test for a shift in mean (bias) of the target distribution across all participants, we use the group-level parameter *shift*. If we find that the 95% HDI does not overlap with zero and is entirely negative, we conclude that repulsion bias was credibly observed across participants (a nonoverlapping positive HDI would be evidence for attraction bias). To quantify swap errors, we followed a similar approach using the group-level parameter *Pswap* (proportion of swapping).

When we test if one condition demonstrated credibly stronger bias or swapping than another condition, we use the HBMM with three levels. The aforementioned group-level shift parameter becomes two separate condition-level parameters, $shift_1$ and $shift_2$, referring to each condition (e.g., trials with 45° color distance use $shift_1$ and trials with 90° color distance use $shift_2$). We compare the posteriors for these condition-level parameters by computing the difference between $shift_1$ and $shift_2$ for every sample and then calculate the 95% HDI of this difference of posteriors (this approach is hereafter referred to as the "within-subject" HBMM). If the resulting HDI does not overlap with 0, we consider this to be a credible difference between conditions (Kruschke, 2014).

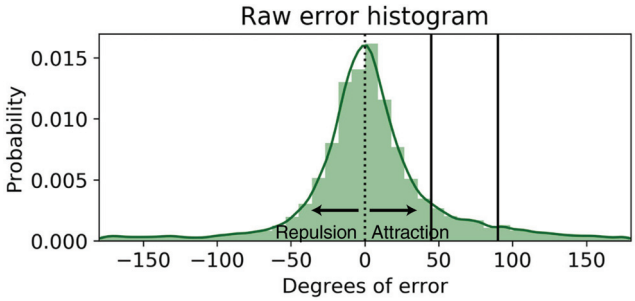
In addition to this preregistered HDI approach to assess whether shift and swap errors were credibly present, we employed exploratory model comparison to assess the contributions of the shift and swap parameters to the model fits. We did this using the widely applicable information criterion (WAIC; Watanabe, 2010), which is computed by estimating how well a model fits the input data while penalizing more complex models. WAIC was chosen because it is fully Bayesian and was previously observed to be more robust than the deviance information criterion (DIC) in similar working memory models (Oberauer et al., 2017). WAIC closely approximates Bayesian cross-validation and is more stable than DIC because variance is separately computed for each sample and then summed, yielding increased stability (Vehtari et al., 2017). To be comparable to AIC or DIC, we report WAIC estimates on the deviance scale, such that the expected log pointwise predictive density for each sample is multiplied by −2. To approximate the uncertainty of WAIC estimates, we calculated the standard error of the difference in WAIC values for each sample (WAIC estimation was performed using the "loo" R package; Vehtari et al., 2019). To assess the contribution of the shift parameter, we compared the (full) HBMM model to the same model without a flexible target mean by examining the difference in WAIC estimates. If the full model demonstrates a smaller WAIC estimate than the nested model then this suggests that the full model better fits the data; that is, the shift parameter explained nontrivial variance in the memory response distribution. We apply the same procedure comparing the full model to the analogous model without a swap parameter to assess the contribution of the swap parameter.

Results and Discussion

Model Results: Overall Shift Errors

We first report overall shift error results before proceeding to shift errors dependent on color distance. Figure 2 depicts the raw histogram of memory response errors across participants, and Figure 3 depicts the group-level (shift) and individual-level estimates for the center (bias) of the target distribution from the hierarchical Bayesian mixture model (HBMM). Table 1 depicts the MAP and 95% highest density interval (HDI) for group-level parameters. The HDI for the shift parameter contained zero; therefore, we did not find credible evidence for shift errors (repulsion or attraction).

Figure 2
Raw Histogram of Responses Across All 50 Participants, Depicting Degrees of Error (Distance Between Reported Color and Actual Color)



Note. Aligned such that the nontarget colors (swap locations) are centered at +45 and +90, depicted as vertical black lines. See the online article for the color version of this figure.

Exploratory model comparison supported the full model compared with the nested model that lacked a flexible target mean, as indicated by a smaller WAIC (full model: 14943.9, nested model: 15261.7; Δ: 317.8, *SE*: 27.8). While model comparison suggested that the shift parameter was an important addition to the model, we lacked evidence for practically meaningful shift errors as indicated by the HDI overlapping with zero. One possibility is that shift errors were not consistent across subjects, such that the additional parameter improved fits for many individual subject estimates but failed to result in a consistent group-level shift parameter. In other words, allowing the target distribution to flexibly shift improved the fit of the model, but the magnitude of this shift was not large or consistent enough to be considered credible.

Shift Errors as a Function of Color Distance

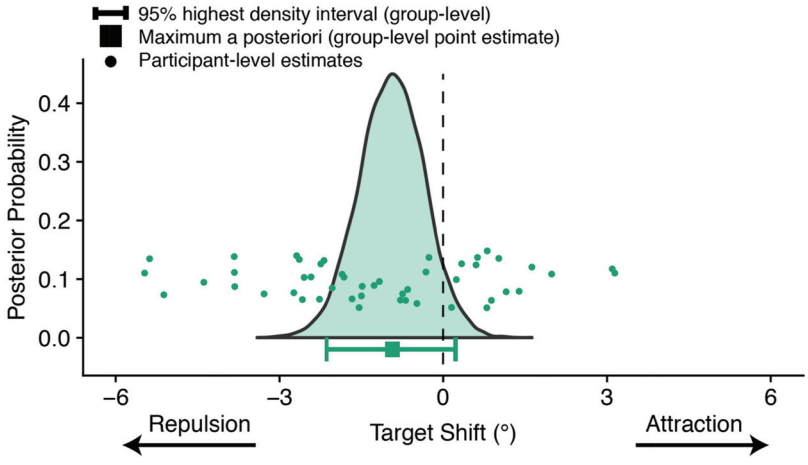
To determine whether shift errors interacted with the color difference between the target and nontarget objects, we fit a within-subject HBMM with color difference (45°, 90°) as the within-subject conditions. We observed no credible shift errors in the 90° (MAP: −.30, HDI: [−1.70, .91]) condition; however, we did observe credible shift errors (repulsion bias) in the 45° condition (MAP:−1.67, HDI: [−2.93, −.42]).

The HDI for the difference in shift posteriors between the two conditions overlapped with zero (MAP: 1.33, HDI: [−.51, 3.05]), so we cannot claim that the 45° color difference was associated with credibly more repulsion than the 90° condition. That said, the presence of a subtle repulsion bias in the 45° color difference is notable in the sense that it demonstrates the capacity for our paradigm to produce shift errors, and the direction of the bias (repulsion) was predicted by the relational representation model (Bae & Luck, 2017; see also Golomb, 2015). In addition to shift errors, there were also no credible differences between conditions for any of the other model parameters (see Table 1 and online supplemental material).

Swap Errors

The HDI for the swap parameter, Pswap, did not contain zero (see Table 1), so we concluded that swap errors were credibly present in Experiment 1. In addition, exploratory model comparison supported the full model compared with the nested model that lacked a swap parameter, as indicated by a smaller WAIC (full model: 14943.9, nested model: 16579.0; Δ: 1635.1, *SE*: 72.2). This demonstrates that swap errors can be observed with real-world objects and are not restricted to paradigms using perceptually similar stimuli (e.g., Chen et al., 2019; Dowd & Golomb, 2019; Golomb et al., 2014).

Figure 3
Split Violin Plot Depicts the Posterior Distribution for the Group-Level Estimate of Shift Errors (Shift) in Experiment 1



Note. The interval underneath represents the 95% HDI, which contained zero, meaning that we did not have credible evidence to support the presence of shift errors (repulsion or attraction). See the online article for the color version of this figure.

Table 1
Group-Level Parameter Estimates for Experiment 1, Including the Maximum a Posteriori (Point Estimate) and the Lower and Upper Bounds of the 95% Highest Density Interval ($HDI_{2.5}$ and $HDI_{97.5}$, Respectively)

Condition	Ptarget	Pswap	Pguess	Shift	SD_{target}	SD_{swap}
All color differences						
MAP	.780	.104	.141	−0.932	23.599	27.446
$HDI_{2.5}$	.719	.062	.088	−2.138	21.431	24.257
$HDI_{97.5}$	.825	.107	.197	0.229	26.769	30.967
45° color difference						
MAP	.785	.086	.126	−1.665	23.739	23.739
$HDI_{2.5}$	.714	.051	.044	−2.930	21.084	21.084
$HDI_{97.5}$	.866	.125	.197	−0.418	28.879	28.879
90° color difference						
MAP	.785	.072	.138	−0.301	24.525	24.525
$HDI_{2.5}$	.728	.050	.079	−1.699	21.848	21.848
$HDI_{97.5}$	.846	.099	.202	0.907	28.752	28.752

Note. Ptarget, Pswap, and Pguess refer to the proportion of target, nontarget, and random guessing responses, respectively. Shift refers to the degrees the target distribution was shifted either towards (attraction; positive values) or away from (repulsion; negative values) the nontarget color. SD_{target} and SD_{swap} reflect the kappa(1) and kappa(2) parameters, converted to degrees of standard deviation. See online supplemental material for non-hierarchical parameter estimates. MAP = maximum a posteriori.

Experiment 2: 3-S Maintenance Duration

It is possible that we did not observe a main effect of shift errors or interaction between shift errors and color distance in Experiment 1 because such memory biases reflect a process that builds during the maintenance interval, and a 1-s working memory delay was too short for this process to be adequately observed. Other experiments have observed attraction/repulsion bias using shorter working memory delays (e.g., Bae & Luck, 2017; Chen et al., 2019; Golomb, 2015), but these studies used simple geometric shapes and spatial or temporal cuing (any of which could explain a shorter time-course for repulsion bias). For Experiment 2, we increased the maintenance duration from 1 s to 3 s. In addition to attraction and/or repulsion bias, we also hypothesized that swap errors would again be present, and might even be larger, with a longer maintenance delay. To preview the results, we observed both repulsion bias and swap errors in this experiment (for both color distances), and we then directly compared the two maintenance durations using a within-subjects manipulation in Experiment 3.

Method

Like Experiment 1, Experiment 2 included a preregistered (https://osf.io/usrxq/?view_only=6d1f0755173f43b3b6d8e8d161dd7fdc) sample size of 50 participants (28 male, 22 female; M = 35.62 years, SD = 8.91).

All participants were recruited using MTurk in the same manner as Experiment 1. Two participants were excluded based on the same preregistered exclusion criteria as Experiment 1. The stimuli, procedure, and analyses for Experiment 2 were identical to Experiment 1 except that the blank interval was increased from 800 ms to 2,800 ms (i.e., working memory delay increased from 1 s to 3 s).

Results and Discussion

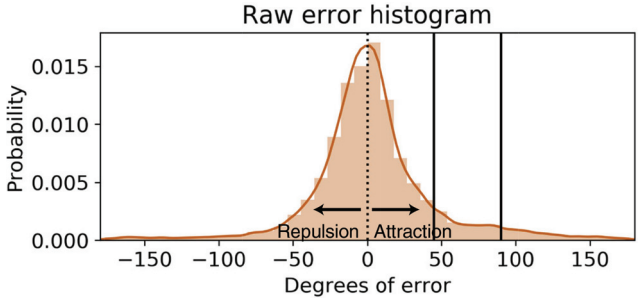
Model Results: Overall Shift Errors

With the 3-s working memory delay, credible repulsion bias was observed where subjects reported a target color biased slightly away from the nontarget color. Figure 4 depicts the raw error histogram across participants, and Figure 5 depicts the group-level (shift) and individual-level estimates for the center of the target distribution. (Note that it may be difficult to visually detect shift errors in the raw histogram because of the opposing push and pull of repulsion bias and swapping.) Table 2 depicts the MAP and 95% HDI for group-level parameters. The HDI for the shift parameter contained only negative values, and exploratory model comparison supported the full model compared with the nested model that lacked a flexible target mean, as indicated by a smaller WAIC (full model: 20761.0, nested model: 21000.4; Δ: 239.4, SE: 29.4).

Shift Errors as a Function of Color Distance

To determine whether shift errors interacted with the color difference between the target and nontarget objects, we fit a within-subject HBMM with color difference as the within-subject conditions. We observed repulsion bias in both the 45° (MAP: −2.78, HDI: [−3.66, −1.62]) condition and the 90° condition (MAP: −.94, HDI: [−1.71, −.03]).

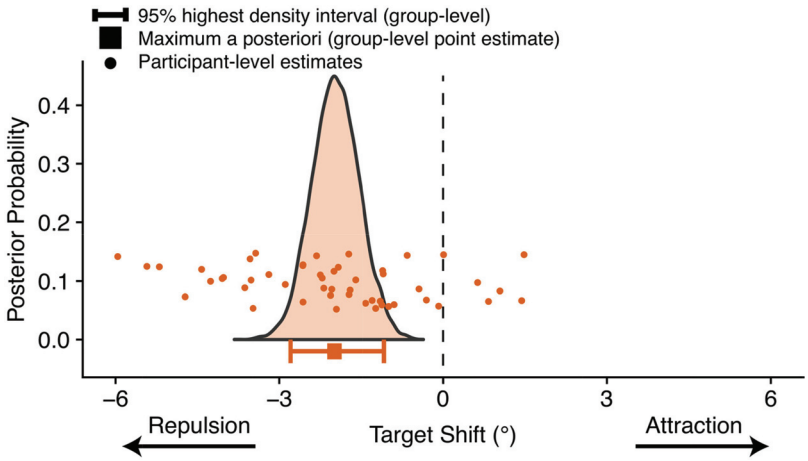
Figure 4
Raw Histogram of Responses Across All 50 Participants, Depicting Degrees of Error (Distance Between Reported Color and Actual Color)



Note. Aligned such that the nontarget colors (swap locations) are centered at +45 and +90, depicted as vertical black lines. See the online article for the color version of this figure.

Figure 5

Split Violin Plot Depicts the Posterior Distribution for the Group-Level Estimate of Shift Errors (Shift) in Experiment 2



Note. The interval underneath represents the 95% HDI, which contained only negative values, meaning that repulsion bias was credibly observed. See the online article for the color version of this figure.

The 45° condition produced credibly stronger repulsion bias than the 90° condition, as indicated by the HDI for the difference in shift posteriors not overlapping with zero (full: MAP: 1.23, HDI: [.01, 2.68]; simplified: MAP: 1.77, HDI: [.43, 3.07]). This suggests that the smaller relative difference in color space led to stronger repulsion bias, in accordance with the relational representation model (Bae & Luck, 2017; see also Golomb, 2015). Even with the relatively larger 90° difference in color space, however, repulsion bias was still observed.

Swap Errors

The HDI for the swap parameter, Pswap, did not contain zero (see Table 2), suggesting that swap errors were again credibly present in Experiment 2. Exploratory model comparison also supported the full model compared with the nested model that lacked a swap parameter, as indicated by a smaller WAIC (full model: 20761.0, nested model: 22888.4; Δ: 2127.4, *SE*: 83.8).

Experiment 3: 1-s Versus 3-s Maintenance Duration

Experiments 1 and 2 suggest that repulsion bias may be modulated by maintenance duration, such that a main effect of repulsion bias was found in our paradigm with a 3-s, but not a 1-s, maintenance duration. In Experiment 3, we test this more directly by recruiting more participants and using a within-subject design where the working memory delay could be either 1 s or 3 s long on a given trial.

Table 2

Group-Level Parameter Estimates for Experiment 2

Condition	Ptarget	Pswap	Pguess	Shift	SD_{target}	SD_{swap}
All color differences						
MAP	.831	.090	.076	−1.994	24.358	33.421
$HDI_{2.5}$	.795	.068	.044	−2.793	21.197	25.511
$HDI_{97.5}$	.867	.112	.115	−1.084	28.179	82.920
45° color difference						
MAP	.856	.064	.075	−2.778	25.101	25.101
$HDI_{2.5}$	.813	.039	.042	−3.661	21.638	21.638
$HDI_{97.5}$	.898	.091	.115	−1.621	31.999	31.999
90° color difference						
MAP	.838	.077	.087	−0.944	23.974	23.974
$HDI_{2.5}$	.794	.057	.047	−1.708	21.321	21.321
$HDI_{97.5}$	.878	.100	.129	−0.026	29.056	29.056

Note. Ptarget, Pswap, and Pguess refer to the proportion of target, nontarget, and random guessing responses, respectively. Shift refers to the degrees the target distribution was shifted either towards (attraction; positive values) or away from (repulsion; negative values) the nontarget color. SD_{target} and SD_{swap} reflect the kappa(1) and kappa(2) parameters (see online supplemental material), converted to degrees of standard deviation. For the separate 45° and 90° color difference models, the standard deviation was shared between the target and swap distributions because of otherwise unreliable parameter estimates (see online supplemental material for parameter estimates with separate precision parameters and for nonhierarchical parameter estimates). MAP = maximum a posteriori; HDI = highest density interval.

In addition, a within-subject design helped to control for the possible confound of participants anticipating the end of the working memory delay. That is, participants might not have evenly focused on maintaining the two representations throughout the working memory delay if they had prior knowledge of how long the delay would last. A within-subject design ensures similar expectation between conditions by randomly altering the length of the working memory delay to either a 1- or 3-s duration.

Method

Experiment 3 included a preregistered (https://osf.io/usrxq/?view_only=6d1f0755173f43b3b6d8e8d161dd7fdc) sample size of 81 participants (39 male, 42 female; M = 40.22 years, SD = 12.19) based on power analyses of Experiments 1 and 2. All participants were recruited using MTurk in the same manner as Experiments 1–2. Four participants were excluded based on the same preregistered exclusion criteria as Experiments 1–2.

The stimuli, procedure, and analyses for Experiment 3 were identical to Experiment 1 except that the working memory delay could be either 1 s or 3 s long (200-ms mask followed by either 800-ms or 2,800-ms blank interval). Each trial condition combination (1s/45° color difference, 1s/90°, 3s/45°, 3s/90°) was presented for 70 trials each (280 total trials), with presentation order randomized for each subject.

Results and Discussion

Model Results: Overall Shift Errors as a Function of Maintenance Duration

Mirroring the results of the first two experiments, when collapsing across color distance, credible repulsion bias was observed in the 3-s condition (MAP: −2.34, HDI: [−3.09, −1.76]) but not the 1-s condition (MAP: −.45, HDI: [′1.01, .22]). Table 3 depicts the MAP and 95% HDI for group-level parameters, Figure 6 depicts the raw error histogram across participants for each condition (as noted above, it may be difficult to visually detect shift errors in the raw histogram), and Figure 7 depicts the group-level (shift) and individual-level estimates for the center of the target distribution.

The within-subject HBMM confirmed that repulsion bias was modulated by maintenance duration, where repulsion bias was credibly larger with the longer working memory delay. There was a credible difference in shift errors between conditions; the HDI for the difference in shift posteriors for the 1-s and 3-s conditions did not overlap with zero (MAP: 2.01, HDI: [1.09, 2.88]).

As was the case in the previous experiments, exploratory model comparison supported the full model compared with the nested model that lacked a flexible target mean, as indicated by a smaller WAIC (full model: 32428.4, nested model: 32936.2; Δ: 507.8, *SE*: 30.6). This provides additional evidence for repulsion bias, as the model provided a better fit when the center of the target distribution was allowed to be biased away from the swap distribution.

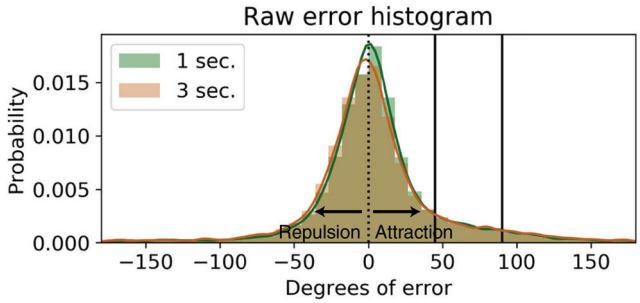
Table 3
Group-Level Parameter Estimates for Experiment 3, Split by Maintenance Duration and Color Distance

Condition	Ptarget	Pswap	Pguess	Shift	SD_{target}	SD_{swap}
3-s maintenance (all color differences)						
MAP	.811	.097	.090	−2.343	23.243	27.064
$HDI_{2.5}$	.775	.079	.063	−3.090	21.152	23.017
$HDI_{97.5}$	.841	.122	.119	−1.761	26.149	33.288
1-s maintenance (all color differences)						
MAP	.836	.092	.075	−0.445	23.196	26.707
$HDI_{2.5}$	.806	.075	.052	−1.008	20.740	22.742
$HDI_{97.5}$	.860	.112	.096	0.222	26.336	33.476
3-s maintenance (45° color diff.)						
MAP	.854	.075	.070	−3.119	24.776	24.771
$HDI_{2.5}$	.814	.044	.051	−3.967	21.708	18.104
$HDI_{97.5}$	.891	.106	.098	−2.010	29.517	66.471
1-s maintenance (45° color diff.)						
MAP	.850	.090	.059	−0.846	23.056	26.904
$HDI_{2.5}$	.814	.064	.017	−1.660	20.413	19.850
$HDI_{97.5}$	.892	.118	.092	0.107	27.189	83.020
3-s maintenance (90° color diff.)						
MAP	.808	.100	.090	−1.389	23.099	24.771
$HDI_{2.5}$	.766	.081	.050	−2.209	21.260	18.104
$HDI_{97.5}$	.845	.128	.128	−0.558	26.057	66.471
1-s maintenance (90° color diff.)						
MAP	.842	.087	.069	−0.061	23.462	26.904
$HDI_{2.5}$	.810	.069	.034	−0.678	20.834	19.850
$HDI_{97.5}$	.878	.108	.102	0.870	27.189	83.020

Note. Ptarget, Pswap, and Pguess refer to the proportion of target, nontarget, and random guessing responses, respectively. Shift refers to the degrees the target distribution was shifted either towards (attraction; positive values) or away from (repulsion; negative values) the nontarget color. SD_{target} and SD_{swap} reflect the kappa(1) and kappa(2) parameters, converted to degrees of standard deviation. For the separate 45° and 90° color difference models, the standard deviation was shared between the target and swap distributions because of otherwise unreliable parameter estimates (see online supplemental material for parameter estimates with separate precision parameters and for nonhierarchical parameter estimates). MAP = maximum a posteriori; HDI = highest density interval.

Figure 6

Raw Histograms per Condition of Responses Across All 81 Participants, Depicting Degrees of Error (Distance Between Reported Color and Actual Color)



Note. Aligned such that the nontarget colors (swap locations) are centered at +45 and +90, depicted as vertical black lines. See the online article for the color version of this figure.

Shift Errors as a Function of Color Distance

To determine whether shift errors interacted with the color difference between the target and nontarget objects, we fit another within-subject HBMM with color difference as the within-subject conditions. When collapsing across maintenance duration, we observed repulsion bias in both the 45° (MAP: −2.10, HDI: [−2.78, −1.39]) condition and the 90° condition (MAP: −.64, HDI: [−1.27, −.01]). The HDI for the difference in shift posteriors did not overlap with zero (MAP: 1.41, HDI: [.51, 2.38]), indicating that the 45° condition produced credibly stronger repulsion bias than the 90° condition, in line with the relational representation model (Bae & Luck, 2017; see also Golomb, 2015) and Experiment 2.

To test for an interaction with maintenance duration, we conducted an exploratory analysis where we separately modeled each maintenance duration with color difference as the within-subject condition (see Table 3). For the 3-s maintenance model, repulsion bias was observed for both the 45° (MAP: −3.12, HDI: [−3.97, −2.01]) and 90° (MAP: −1.39, HDI: [−2.21, −.56]) color difference conditions, with the 45° condition producing credibly stronger repulsion bias than the 90° condition (MAP: −1.55, HDI: [−2.91, −.37]). For the 1-s maintenance model, credible repulsion bias was not observed for either the 45° (MAP: −.85, HDI: [−1.66, .11]) or the 90° (MAP: −.06, HDI: [−.68, .87]) color difference condition, with no credible difference between conditions (MAP: −.89, HDI: [−2.03, .31]). This suggests that the overall stronger repulsion bias for the 45° color difference was likely driven by the 3-s maintenance trials.

Swap Errors

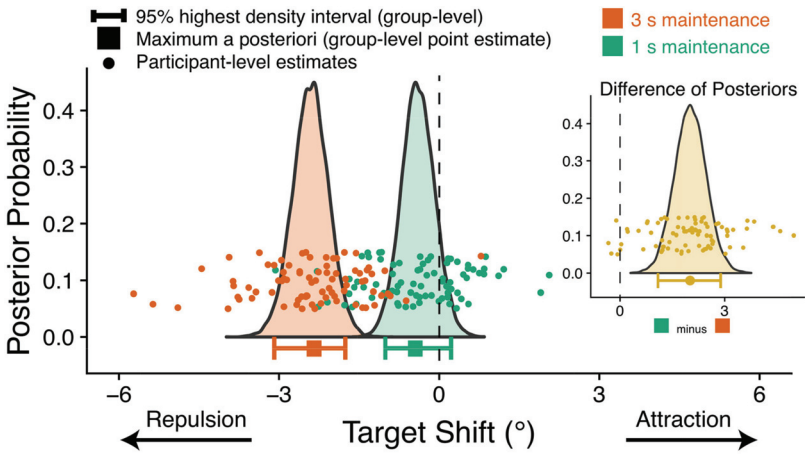
For both maintenance duration conditions, the HDI for the swap parameter, Pswap, did not contain zero (see Table 3), indicating that swap errors were present regardless of the maintenance duration. Exploratory model comparison further supports this claim, as the full model outperformed the nested model that lacked a swap parameter (full model: 32428.4, nested model: 37028.1; Δ: 4599.7, *SE*: 122.8). The HDI for the difference in Pswap posteriors contained zero (MAP: .008, HDI: [−.020, .038]), indicating no credible difference in the proportion of swap errors between the 1-s and 3-s conditions.

Memory Performance

In an exploratory analysis, we compared the target parameter, Ptarget, between maintenance duration conditions to test whether the longer maintenance duration led to overall worse performance. The HDI for the difference in Ptarget posteriors overlapped with zero (MAP: .019, HDI: [−.018, .068]), indicating no credible difference in the proportion of target responses between maintenance duration conditions. The Pguess parameter, reflecting random

Figure 7

Split Violin Plots Depict the Posterior Distributions for the Group-Level Estimates of Shift Errors (Shift) in Experiment 3, Split by Maintenance Duration



Note. The inset depicts the difference of posteriors, which did not contain zero, meaning that repulsion bias was credibly larger for the longer maintenance duration. See the online article for the color version of this figure.

guessing, also showed no credible difference between conditions (MAP: .015, HDI: [−.017, .054]). There was also no credible difference between conditions for SD_{target} (MAP: 3.21, HDI: [−107.78, 107.15]) and SD_{swap} (MAP: −7.37, HDI: [−125.53, 113.68]). Overall, these exploratory comparisons suggest that the stronger repulsion bias for the 3 s maintenance duration was unlikely to be driven by a difference in overall memory performance between conditions.

Experiment 4: Filler Versus No-Filler Task

The results of Experiments 1–3 indicated that repulsion bias is stronger with a longer working memory delay. The existence of a repulsion bias in this paradigm suggests that repulsion bias is not necessarily retinotopically specific and can be attributed to a mnemonic, not a perceptual, explanation. We also demonstrated swap errors in all experiments, suggesting that swap errors can be observed with perceptually distinct, real-world objects. In Experiment 4, we aimed to better understand why increasing maintenance duration leads to greater repulsion bias.

One hypothesis is that repulsion bias could be explained by an active maintenance process; that is, it could be that repulsion bias occurs as the result of multiple representations competing for attention in working memory. The longer time spent actively attending to representations during maintenance, the more these representations may systematically repel from each other to produce less interitem confusion. This explanation would be in line with the theory of biased competition, where representations compete for cortical activity, influenced both by sensory activity and top-down attentional biases (Desimone & Duncan, 1995).

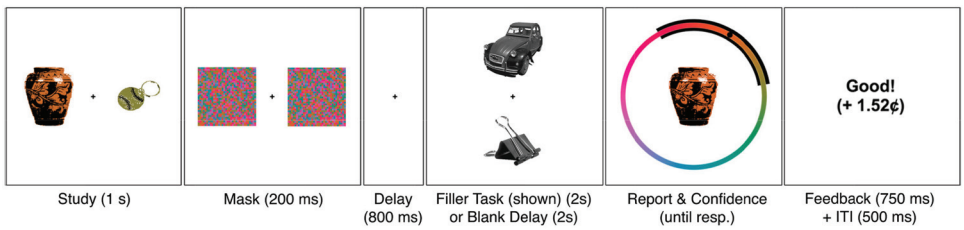
An alternative hypothesis is that repulsion occurs as a function of memory degradation. If longer memory delays lead to poorer quality representations (e.g., due to passive memory degradation or increased contextual interference; Barrouillet et al., 2012; Davis & Zhong, 2017; Oberauer & Lewandowsky, 2008), this may prompt the memory system to prioritize other sources of information at recall, such as relational information, resulting in increased repulsion bias. That memory degradation might result in increased mnemonic bias follows predictions from an ideal (or optimal) observer model, that attraction and repulsion biases might be adaptive because they reflect the Bayesian procedure of combining uncertain item-level information (i.e., representation of the target item's color) with other available information (Brady et al., 2018; Chunharas et al., 2019; Geisler, 2011; Hemmer & Steyvers, 2009; Honig et al., 2020; Huttenlocher et al., 2000). Here the other available information (often referred to as group-level information or priors) would be about other items in the display, including relational information like feature similarity or relative distance in feature space between memory items, and the idea is that this group-level information is weighted more heavily when the item-level information is less certain; thus, repulsion bias would strengthen as the quality of memory representations weakens over time.

In Experiment 4 we added a filler task presented during the blank interval, because it leads to two opposite predictions according to the above active maintenance and memory degradation accounts (see Figure 8). Specifically, in Experiment 4, half of the trials had a blank delay of 2,800 ms and the other half of trials included a filler task during the delay period. In this filler task condition, the trials had a blank delay of 800 ms followed by 2 s to perform a size comparison task between two unique grayscale objects (which object has the larger real-world size). If repulsion bias reflects an active process during maintenance, then a filler task should interfere with active attention to internal representations and result in decreased repulsion bias. If repulsion bias occurs as a function of memory degradation, then a filler task should worsen the quality of memory representations and hence increase repulsion bias.

Note that our theoretical accounts operate under the assumption that memory items were sufficiently encoded and maintained (actively or passively). In the extreme example where a participant is hardly able to recognize items, let alone remember their exact color, it may be optimal for the memory system to ignore subtle color differences and instead prioritize a gist-based representation (e.g., average color of all memory items, which might actually produce attraction bias) or to prioritize one of the two items while discarding the other (in the hopes that the discarded item will not be tested). We designed our experiments with the aim to provide enough time for participants to encode and report the memory items such that always responding around the average color, or prioritizing one item and discarding the other item, would be suboptimal strategies. To preview our results, we did not find evidence for attraction bias, and the proportion of target responses averaged over 70% in both conditions, suggesting that participants sufficiently encoded the memory items (although it is difficult to

Figure 8

Example Trial Sequence for Experiment 4



Note. The trial procedure was identical to Experiment 2 except that half of the trials involved a 2-s size comparison task during the delay, where the subject was instructed to click on the grayscale object with the larger real-world size (in this example, the subject would click the car). The absolute time between encoding and test was 3 s, regardless of whether the trial contained the filler task or not. See the online article for the color version of this figure.

wholly rule out these strategies based on our results). Also note that the bias parameter is tied to the target distribution, such that repulsion/attraction bias is only influencing trials where the item is thought to be successfully maintained (according to the HBMM).

Finally, if we find no difference in repulsion bias between the filler and no-filler conditions, it is possible that mnemonic repulsion bias could be explained by a third hypothesis, a simple temporal decay explanation. That is, whereas the active account predicts reduced repulsion in the filler condition (due to lack of active attention) and the degradation account predicts increased repulsion in the filler condition (due to decreased memory quality), the temporal decay account suggests that the effect might be driven simply by the passage of time, resulting in similar repulsion regardless of the filler task. Temporal decay is a major factor in forgetting (Barrouillet & Camos, 2009; Barrouillet et al., 2012; Pertzov et al., 2013; Vergauwe et al., 2009); hence, it might be possible that the longer representations spend in maintenance, the greater the repulsion bias. If repulsion bias can be explained by temporal decay, then it should not matter if there is a filler task or not during the working memory delay, as long as the absolute time between encoding and retrieval is the same between conditions. However, temporal decay is a relatively unclear mechanism. For instance, temporal decay is correlated with several other cognitive variables including contextual interference, and after controlling for these variables it has been suggested that forgetting in working memory does not depend at all on temporal decay (Lewandowsky & Oberauer, 2009; Oberauer & Kliegl, 2006; Oberauer & Lewandowsky, 2008). Thus, a lack of credible difference between the filler and no-filler conditions would not be as conclusive a result from a theoretical perspective.

Method

Experiment 4 included a preregistered (https://osf.io/usrxq/?view_only=6d1f0755173f43b3b6d8e8d161dd7fdc) sample size of 81 participants (46 male, 35 female; M = 38.28 years, SD = 12.58). All participants were recruited using MTurk in the same manner as Experiments 1–3. Thirty-four participants were excluded based on preregistered exclusion criteria: Four participants were excluded based on the same criteria as Experiments 1–3 (>.5 guessing proportion based on a basic mixture model), and 30 participants were excluded for not correctly performing the filler task (accuracy < 75%; note that because the filler task did not require input, most of these excluded participants appeared to ignore the filler task entirely).

The stimuli, procedure, and analyses for Experiment 4 were identical to Experiment 2 except that half the trials contained an intervening filler task during the delay (see Figure 8). Each trial condition combination (filler/45° color difference, no-filler/45°, filler/90°, no-filler/90°) was presented for 70 trials each (280 total trials), with presentation order randomized for each subject. The filler task was a size comparison task consisting of two unique real-world objects presented above and below the fixation cross (150 pixels away from fixation cross). The objects were the same physical size as the memory items (each object was 200 × 200 pixels), and the task was to indicate which object was of the larger *real-world* size. These objects were grayscale and drawn from a separate stimulus set than the objects presented during encoding. Objects were drawn from the Big and Small Objects dataset (Konkle & Oliva, 2012) and the two displayed objects always consisted of one "small" and one "big" object.

We chose this object size comparison task as the filler task because a task requiring subjects to meaningfully process other real-world objects should produce substantial interference with the real-world objects being held in working memory (Craik, 2014), allowing us to test whether repulsion bias depends on active maintenance of memory representations. We positioned the size comparison stimuli in different spatial locations than the initially encoded objects because we were not interested in interference from sensory memory or overlapping retinotopic information.

Results and Discussion

Model Results: Overall Shift Errors as a Function of Filler Task

Mirroring the results of Experiments 2 and 3, for the 3-s maintenance duration, no-filler condition, the HDI for the shift parameter contained only negative values (MAP: −1.60, HDI: [−2.26, −.93]). For the 3-s filler condition, however, the HDI contained zero (MAP: −.13, HDI: [−.96, .76]). Credible repulsion bias was therefore observed in the no-filler condition but not in the filler condition. Table 4 depicts the MAP and 95% HDI for group-level parameters, Figure 9 depicts the raw error histogram across

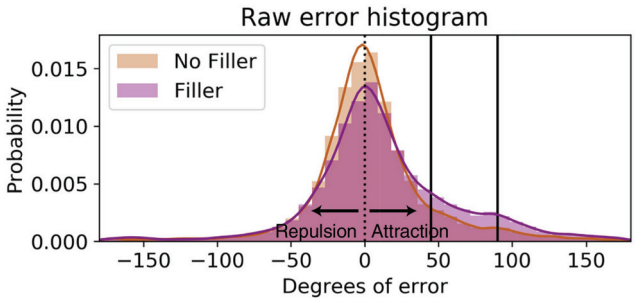
Table 4
Group-Level Parameter Estimates for Experiment 4, Split by Condition

Condition	Ptarget	Pswap	Pguess	Shift	SD_{target}	SD_{swap}
No filler						
MAP	.850	.089	.061	−1.600	24.252	29.477
$HDI_{2.5}$	.815	.069	.031	−2.259	22.160	24.737
$HDI_{97.5}$	.881	.117	.089	−0.930	27.535	47.416
Filler						
MAP	.720	.197	.085	−0.132	27.220	31.302
$HDI_{2.5}$	.687	.170	.063	−0.955	24.445	27.300
$HDI_{97.5}$	.749	.224	.110	0.759	31.652	38.941

Note. Ptarget, Pswap, and Pguess refer to the proportion of target, nontarget, and random guessing responses, respectively. Shift refers to the degrees the target distribution was shifted either towards (attraction; positive values) or away from (repulsion; negative values) the nontarget color. SD_{target} and SD_{swap} reflect the kappa(1) and kappa(2) parameters, converted to degrees of standard deviation. See online supplemental material for nonhierarchical parameter estimates. MAP = maximum a posteriori; HDI = highest density interval.

Figure 9

Raw Histograms per Condition of Responses Across All 81 Participants, Depicting Degrees of Error (Distance Between Reported Color and Actual Color)



Note. Aligned such that the nontarget colors (swap locations) are centered at +45 and +90, depicted as vertical black lines. See the online article for the color version of this figure.

participants for each condition (as noted above, it may be difficult to visually detect shift errors in the raw histogram), and Figure 10 depicts the group-level (shift) and individual-level estimates for the center of the target distribution.

The within-subject HBMM found that repulsion bias was stronger on no-filler trials than filler trials, as indicated by a credible difference in shift errors between conditions. The HDI for the difference in shift posteriors did not overlap with zero (MAP: 1.46, HDI: [.46, 2.63]). This pattern of results supports the idea that repulsion bias reflects an active maintenance process, which the filler task interferes with.

As was the case in the previous experiments, exploratory model comparison supported the full model compared with the nested model that lacked a flexible target mean, as indicated by a smaller WAIC (full model: 37494.8, nested model: 37771.0; Δ: 276.2, *SE*: 23.4). This provides additional support that repulsion bias was credible, as the model provided a better fit when the center of the target distribution was allowed to be biased away from the swap distribution.

Shift Errors as a Function of Color Distance

As in previous experiments, we fit another within-subject HBMM with color difference as the within-subject conditions. When collapsing across filler and no-filler conditions, we observed repulsion bias in the 45° (MAP: −1.95, HDI: [−2.84, −1.22]) condition but not the 90° condition (MAP: −.06, HDI: [−.82, .74]). The HDI for the difference in shift posteriors did not overlap with zero (MAP: 1.98, HDI: [.83, 3.07]), indicating that the 45° condition produced credibly stronger repulsion bias than the 90° condition, in line with the relational representation model (Bae & Luck, 2017; see also Golomb, 2015) and Experiments 2 and 3.

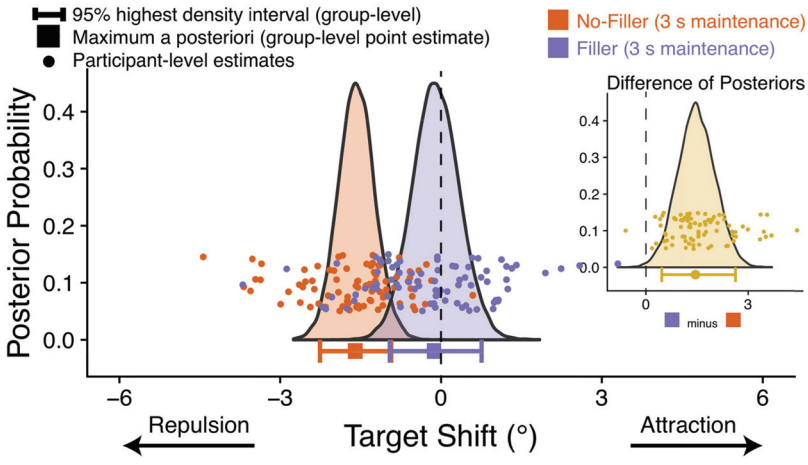
To test for an interaction with no-filler/filler task, we conducted an exploratory analysis where we separately modeled filler and no-filler trials, with color difference as the within-subject condition (see Supplementary Tables 9–10). For the no-filler model, credible repulsion bias was observed for the 45° condition only, with the 45° condition producing credibly stronger repulsion bias than the 90° condition. For the filler model, credible repulsion bias was not observed for either color difference condition, with no credible difference between conditions. This suggests that the overall stronger repulsion bias for the 45° color difference was likely driven by the no-filler trials.

Swap Errors

Swap errors (mistakenly reporting the nontarget color) increased in the presence of a filler task, as indicated by a credible difference in Pswap between conditions (MAP: .104, HDI: [.069, .141]). While the proportion of swap errors was larger on filler trials compared to no-filler trials, swap errors were credibly observed

Figure 10

Split Violin Plots Depict the Posterior Distributions for the Group-Level Estimates of Shift Errors (Shift) in Experiment 4, Split by Condition (No-Filler or Filler)



Note. The inset depicts the difference of posteriors, which did not contain zero, meaning that repulsion bias was credibly stronger for trials that did not contain a filler task. See the online article for the color version of this figure.

in both conditions (see Table 4), as indicated by the HDI for both Pswap1 (group-level swap parameter for no-filler condition) and Pswap2 (group-level swap parameter for filler condition) not containing zero. Exploratory model comparison supported that swap errors were credibly observed, as the full model outperformed the nested model that lacked a swap parameter (full model: 37494.8, nested model: 43921.6; Δ: 6426.8, *SE*: 151.4).

Memory Performance

In an exploratory analysis, we compared model parameter estimates between no-filler/filler conditions to test whether the filler task led to overall worse performance (see Table 4). As expected, filler task trials led to worsened memory performance as indicated by a decreased proportion of target responses. The HDI for the difference in Ptarget posteriors did not overlap with zero (MAP: .130, HDI: [.086, .178]), indicating a credible difference in the proportion of target responses between conditions. Meanwhile, the Pguess parameter, reflecting random guessing, showed no credible difference between conditions (MAP: .024, HDI: [−.011, .062]), though there was an increase in swap errors, as reported above. There was no credible difference between conditions for either *SD* measure: SD_{target} (MAP: 72.12, HDI: [−23.38, 154.33]) and SD_{swap} (MAP: 19.73, HDI: [−116.50, 110.25]).

Summary

Stronger repulsion bias was observed on trials without a filler task compared with trials with a filler task, supporting the idea that repulsion bias reflected an active maintenance process. Combined with the results of Experiment 3, repulsion bias seems to occur as the result of multiple memory representations competing for attention. Conversely, swap errors were observed more often on trials with a filler task than trials without a filler task, following the consensus that spatial attention is crucial for object-feature integrity (e.g., Dowd & Golomb, 2019; Emrich & Ferber, 2012; Robertson, 2003; Treisman & Schmidt, 1982; Vul & Rich, 2010).

General Discussion

The main contribution of this article is that repulsion bias strengthened with longer working memory delays, but only when items were actively maintained. Repulsion bias reflects the subtle misremembering of a target memory item as more dissimilar to a reference point than it is in reality, likely in an attempt to better differentiate items by the memory system (Bae & Luck, 2017; Chunharas et al., 2019; Golomb, 2015). The process underlying repulsion bias was found to occur during maintenance in a nonretinotopically specific experimental design, suggesting that repulsion bias can occur mnemonically (that is, in the absence of a perceptual explanation). Moreover, Experiment 4 revealed that a filler task during the working memory delay could disrupt the effect, suggesting that this mnemonic repulsion bias is an active process. Below we discuss the neural mechanisms and psychological theories that could potentially support this finding of active mnemonic repulsion bias.

Neural Mechanisms of Repulsion Bias

Repulsion bias (perceptual and/or mnemonic) has generally been explained by the underlying neural mechanisms of lateral inhibition or optimal gain. According to lateral inhibition explanations, object features (like color) may be represented in a map-like way such that neighboring neurons code neighboring parts of feature space, and lateral inhibitory connections between them help sharpen feature representations. Thus, if the features of an attended working memory object are similar to the features of another working memory object, such that their representations are coded with nearby neurons, these neurons may inhibit each other. As a result, neurons representing this similar region of feature space become relatively suppressed, which effectively results in both feature representations becoming biased away from each other, and repulsion bias is observed (Johnson et al., 2009; Wei et al., 2012).

Such theories of lateral inhibition have sometimes been explained within the framework of dynamic field theory (or more generally, continuous-attractor neural network models: Amari, 1977; Wilson & Cowan, 1972), where recurrent interactions explain how neural projections are capable of sustaining themselves even in the absence of sustained perceptual input, providing a possible neural explanation for active working memory maintenance (Johnson et al., 2014; Schutte & Spencer, 2009; Simmering & Spencer, 2008; Simmering et al., 2006; Spencer et al., 2007). Specifically, dynamic field theory has been used to explain evidence of delay-dependent repulsion bias in spatial working memory tasks (e.g., Simmering & Spencer, 2008) and therefore may generalize to our present observations of delay-dependent repulsion bias in nonspatial visual working memory.

Whereas theories of lateral inhibition (specifically dynamic field theory) support an active process of visual working memory maintenance, the optimal gain account makes no distinction between active and passive processing. Optimal gain theory asserts that when a target and nontarget are highly similar, the optimal behavior is to increase the salience of the target relative to the nontarget. From a single neuron perspective, this means that enhancing the response of neurons that are tuned slightly away from the target (in the direction away from the nontarget) can maximize the signal-to-noise ratio between the target and the nontarget, enabling better discrimination (Navalpakkam & Itti, 2007; Scolari & Serences, 2009).

Optimal gain theory could support the active maintenance account by suggesting that focused attention to the memory items during a delay allows for increasingly more precise tuning to maximize the signal-to-noise ratio. But optimal gain theory could also support a memory degradation account because the optimal behavior may be that tuning should become more exaggerated when there is interrupted attention, such that increased enhancement of neurons tuned away from the nontarget item prevents interitem confusion during a lapse of attention. Optimal gain may occur in combination with lateral inhibition to support behavioral evidence of repulsion bias.

Independent of the lateral inhibition versus optimal gain mechanisms, a recent study using human EEG observed neural evidence consistent with active memory mechanisms. Sutterer et al. (2019) found that the selectivity of population-level tuning functions decreased with two items in maintenance compared with one item. They then used data-driven simulations to conclude that a working memory model where two items can be simultaneously active is

better supported as opposed to models where two items are maintained by rapidly switching between single-item active states or by keeping one item in the focus of attention while others are relegated to a passive long-term memory store. If multiple memory items can be concurrently active in working memory then this would be consistent with the above active maintenance mechanisms and could suggest that neuroimaging investigations of repulsion bias might allow us to further explore the "active" nature of the active maintenance account.

Psychological Theories of Repulsion Bias

Our findings also speak to several psychological theories of repulsion bias, including a low-level perceptual account, theories of working memory interdependence, and an ideal observer model. These psychological theories could be implemented by the above neurobiological accounts of repulsion bias; the psychological theories are not meant to be in opposition to the neural explanations. We first rule out a purely perceptual account, concluding that repulsion bias observed across our experiments reflected a mnemonic phenomenon. We then consider how some canonical theories of working memory that assume that memory items are stored independently are inconsistent with our results. Finally, we discuss whether mnemonic repulsion bias might reflect Bayesian-like behavior in accordance with an ideal observer model.

Perceptual Account

Our first goal was to test a low-level perceptual account for repulsion bias. As explained in the Introduction, repulsion bias can be observed in the absence of mnemonic processing in the case of the direction and tilt illusions (e.g., Gibson & Radner, 1937; Wiese & Wenderoth, 2007). These illusions can then be extended to working memory designs in the case of direction aftereffects and tilt aftereffects, where visual adaptation induces retinotopically specific perceptual biases on subsequent visual input (e.g., Gibson, 1937; Hiris & Blake, 1996; Wenderoth & Johnstone, 1988; Wenderoth & Wiese, 2008; Wiese & Wenderoth, 2007). An open question was whether such perceptual mechanisms might account for the repulsion biases observed in visual working memory experiments.

Our experiments are the first to demonstrate repulsion bias using a nonretinotopically specific working memory design, which rules out a purely perceptual explanation. Moreover, we observed differences in repulsion bias between conditions with identical layout and encoding demands, that only differed in terms of the maintenance process. The mnemonic repulsion bias observed in the present experiments is thus distinct from the above observations of perceptual repulsion bias. As such, the psychological theories discussed in more detail below focus on mnemonic, not perceptual, interactions.

Working Memory Interdependence

Several influential models of visual working memory, including slot models (e.g., Luck & Vogel, 1997; Zhang & Luck, 2008) and resource models (e.g., Bays & Husain, 2008), contend that memory items are stored independently. More recently, these models have been disputed in favor of nonindependent working memory storage (e.g., Brady & Alvarez, 2015; Johnson et al., 2014; Oberauer & Lin, 2017). For instance, statistical regularities in regard to prior stimulus distributions and spatial context can bias working memory reports (Huang & Sekuler, 2010; Jiang et al., 2000). Hypothetically, humans could have an independent memory system where biases occur at decision-making, but our results support the interdependence of memory items because we observed biases in behavior dependent on whether items were actively attended to during this interval. This does not preclude additional biases potentially being involved at decision-making.

Ideal Observer Model

The ideal observer model is a general theoretical framework (Geisler, 2011) that has been applied to mnemonic repulsion bias and intuits that subjects adaptively combine group-level (or gist) information (e.g., color of nontarget items, ensemble statistics, spatial context, or learned "priors") with item-level information about the target object (Brady et al., 2018; Hemmer & Steyvers, 2009; Huttenlocher et al., 2000) to optimize performance, given the limited capacity of visual working memory (Luck & Vogel, 1997; Zhang & Luck, 2008). The ideal observer model could provide a parsimonious explanation with the conclusion of Chunharas et al. (2019)—that repulsion bias is an adaptive process that is stronger with shorter encoding durations and longer maintenance durations—if we consider the relational information (also known as feature similarity) of memory items to be an important group-level feature. Shorter encoding durations and longer maintenance durations should both weaken (item-level) memory strength, and it was in these conditions that Chunharas et al. (2019) observed the strongest repulsion bias (i.e., relational information was relied on by the memory system more in these task conditions).

At first glance, our current results are not consistent with this interpretation of the ideal observer model. In Experiment 4, we observed that repulsion bias was stronger in the condition where the item-level evidence (memory strength as indicated by an increased proportion of target responses) was also stronger. Further, when we used confidence range reports as a proxy for memory strength, we did not observe any association between confidence and the magnitude of repulsion bias (although the confidence reports may not have been sensitive enough to detect a difference, see the online supplemental material), whereas the ideal observer model would predict that highly confident memory reports should be associated with weaker repulsion bias (as in Honig et al., 2020, where they demonstrate how an ideal observer model can explain how participant uncertainty correlated with attraction bias).

If we consider a few restrictions on the ideal observer model, however, then this theory could still provide a valid explanation for active mnemonic repulsion bias. In other words, perhaps relational information might only be relied on under certain circumstances; for instance, when working memory items are (a) sufficiently encoded, (b) easily confusable or not easily individuated, and (c) actively attended. If these conditions are all met, then weakened memory strength should be associated with increased repulsion bias, in line with the ideal observer model. Indeed, in an exploratory analysis where we only included trials where responses were reasonably correct (<30 ° away from the target color), there was a significant negative correlation between repulsion bias and confidence (see the online supplemental material). However, if any of those three conditions are not met, then a more suitable prior to rely on might be gist information

(also known as ensemble statistic, average feature), in which case repulsion errors would be less likely to be observed. As an example, given two similar blue-green objects that are sufficiently encoded and actively attended, active mnemonic processing separates the colors away from each other to reduce confusability (repulsion bias); however, if the two blue-green objects are not sufficiently encoded or maintained, you may only remember "cool" colors and hence select a medium blue-green color (attraction bias).

Additional Considerations

While some prior studies have used a nontarget item as a reference point (e.g., Chunharas et al., 2019; Czoschke et al., 2020), several other studies have used hierarchical properties like the ensemble statistic (e.g., Brady & Alvarez, 2011) or the average color observed across all past trials (e.g., Huang & Sekuler, 2010). We note that our results could differ depending on what property the memory system uses as the reference point for repulsion bias because hierarchical properties may be automatically computed and take up space in memory independently from concurrently maintained memory items (for discussion see Brady et al., 2011). Results could also differ if more than two items need to be maintained or if items must be transferred into long-term memory.

Another interesting question is whether the swap distribution is biased similarly to the target distribution. In theory, a swap represents object-feature misbinding such that the participant may be reporting the successfully maintained nontarget item. In this case, we would expect to observe repulsion bias for the swapped nontarget in addition to our present observation of repulsion bias for the target item (i.e., both items repulsed away from each other during maintenance). On the other hand, it is possible that swap trials indicate that maintenance was not as successful as trials where items were correctly reported, and because we conclude that successful encoding and active attention are necessary components to repulsion bias, perhaps such a bias is not as strong or is not present for swap trials. It is difficult to test for a bias in the swap distribution with the present data due to an insufficient number of swap trials, but future work could explore this research question by prompting color reports for both memory items and/or manipulating the paradigm to encourage more swap errors.

It is also important to consider that the present experiments are the first demonstration of repulsion bias using real-world objects. One could have argued that repulsion bias would only be expected if objects shared similar perceptual features (e.g., memory items are two similarly colored squares). Our observation of credible repulsion bias using perceptually distinct, real-world objects suggests that repulsion bias is robust to more realistic situations and to a large assortment of visual stimuli.

Context of the Research

This research stems from work in visual working memory that observed attraction and repulsion bias dependent on interitem similarity (e.g., Bae & Luck, 2017; Golomb, 2015) and work in the perception domain (specifically tilt and direction aftereffects) that elicited similar attraction and repulsion biases (e.g., Gibson, 1937; Hiris & Blake, 1996; Wenderoth & Johnstone, 1988; Wenderoth & Wiese, 2008; Wiese & Wenderoth, 2007). Recent articles from our research group also observed repulsion and attraction biases in long-term memory using real-world stimuli (Scotti et al., 2021) as well as biased feature perception during dynamic spatial attention contexts such as attentional capture (Chen et al., 2019) and remapping across eye movements (Golomb et al., 2014). The similarities observed across domains prompted theoretical discussion regarding whether similar mechanisms account for these phenomena, and more generally, the functional role and implications of memory biases for models of working memory. There are several directions for future research related to the present work, including behavioral investigations to understand the contexts where repulsion bias is observed and neuroimaging investigations to potentially observe an active representational bias without reliance on behavioral input.

Conclusions

The present results raise important questions regarding the nature of working memory maintenance. We first demonstrate that repulsion bias in visual working memory can be observed in the absence of any perceptual explanation. We then report the somewhat counterintuitive idea that improved attention to working memory items during maintenance can result in stronger repulsion bias. This observation that repulsion bias in working memory reflects an active process ongoing during maintenance supports theories of working memory that assert that representations are interdependent (e.g., Brady & Alvarez, 2015; Oberauer & Lin, 2017), adding that these dependencies can produce systematic biases as the result of competing mnemonic representations.

References

Amari, S. I. (1977). Dynamics of pattern formation in lateral-inhibition type neural fields. *Biological Cybernetics*, *27*(2), 77–87. https://doi.org/10.1007/BF00337259

Awh, E., & Jonides, J. (2001). Overlapping mechanisms of attention and spatial working memory. *Trends in Cognitive Sciences*, *5*(3), 119–126. https://doi.org/10.1016/s1364-6613(00)01593-x

Bae, G.-Y., & Luck, S. J. (2017). Interactions between visual working memory representations. *Attention, Perception & Psychophysics*, *79*(8), 2376–2395. https://doi.org/10.3758/s13414-017-1404-8

Bae, G.-Y., Olkkonen, M., Allred, S. R., & Flombaum, J. I. (2015). Why some colors appear more memorable than others: A model combining categories and particulars in color working memory. *Journal of Experimental Psychology: General*, *144*(4), 744–763. https://doi.org/10.1037/xge0000076

Barrouillet, P., & Camos, V. (2009). Interference: Unique source of forgetting in working memory? *Trends in Cognitive Sciences*, *13*(4), 145–146. https://doi.org/10.1016/j.tics.2009.01.002

Barrouillet, P., De Paepe, A., & Langerock, N. (2012). Time causes forgetting from working memory. *Psychonomic Bulletin & Review*, *19*(1), 87–92. https://doi.org/10.3758/s13423-011-0192-8

Bays, P. M., & Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science*, *321*(5890), 851–854. https://doi.org/10.1126/science.1158023

Bays, P. M., Catalao, R. F., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision*, *9*(10), 7. https://doi.org/10.1167/9.10.7

Bays, P. M., Wu, E. Y., & Husain, M. (2011). Storage and binding of object features in visual working memory. *Neuropsychologia*, *49*(6), 1622–1631. https://doi.org/10.1016/j.neuropsychologia.2010.12.023

Blakemore, C., Carpenter, R. H., & Georgeson, M. A. (1970). Lateral inhibition between orientation detectors in the human visual system. *Nature*, *228*(5266), 37–39. https://doi.org/10.1038/228037a0

Brady, T., & Alvarez, G. (2011). Hierarchical encoding in visual working memory: Ensemble statistics bias memory for individual items. *Psychological Science*, *22*(3), 384–392. https://doi.org/10.1177/0956797610397956

Brady, T., & Alvarez, G. (2015). Contextual effects in visual working memory reveal hierarchically structured memory representations. *Journal of Vision*, *15*(15), 6. https://doi.org/10.1167/15.15.6

Brady, T. F., Konkle, T., & Alvarez, G. A. (2011). A review of visual memory capacity: Beyond individual items and toward structured representations. *Journal of Vision*, *11*(5), 4.

Brady, T., & Tenenbaum, J. (2013). A probabilistic model of visual working memory: Incorporating higher order regularities into working memory capacity estimates. *Psychological Review*, *120*(1), 85–109. https://doi.org/10.1037/a0030779

Brady, T., Konkle, T., Alvarez, G., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(38), 14325–14329. https://doi.org/10.1073/pnas.0803390105

Brady, T., Schacter, D., & Alvarez, G. (2018). The adaptive nature of false memories is revealed by gist-based distortion of true memories. *PsyArXiv*. https://doi.org/10.31234/osf.io/zeg95

Carpenter, A. C., & Schacter, D. L. (2017). Flexible retrieval: When true inferences produce false memories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(3), 335–349. https://doi.org/10.1037/xlm0000340

Chen, J., Leber, A. B., & Golomb, J. D. (2019). Attentional capture alters feature perception. *Journal of Experimental Psychology: Human Perception and Performance*, *45*(11), 1443–1454. https://doi.org/10.1037/xhp0000681

Chunharas, C., Rademaker, R. L., Brady, T. F., & Serences, J., (2019). Adaptive memory distortion in visual working memory. *PsyArXiv*. https://doi.org/10.31234/osf.io/e3m5a

Craik, F. I. (2014). Effects of distraction on memory and cognition: A commentary. *Frontiers in Psychology*, *5*, 841. https://doi.org/10.3389/fpsyg.2014.00841

Czoschke, S., Peters, B., Rahm, B., Kaiser, J., & Bledowski, C. (2020). Visual objects interact differently during encoding and memory maintenance. *Attention Perception & Psychophysics*, *82*(3), 1241–1257. https://doi.org/10.3758/s13414-019-01861-x

Davis, R. L., & Zhong, Y. (2017). The biology of forgetting—A perspective. *Neuron*, *95*(3), 490–503. https://doi.org/10.1016/j.neuron.2017.05.039

de Fockert, J., & Wolfenstein, C. (2009). Rapid extraction of mean identity from sets of faces. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *62*(9), 1716–1722. https://doi.org/10.1080/17470210902811249

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*(1), 193–222. https://doi.org/10.1146/annurev.ne.18.030195.001205

Dowd, E. W., & Golomb, J. D. (2019). Object-feature binding survives dynamic shifts of spatial attention. *Psychological Science*, *30*(3), 343–361. https://doi.org/10.1177/0956797618818481

Emrich, S. M., & Ferber, S. (2012). Competition increases binding errors in visual working memory. *Journal of Vision*, *12*(4), 12. https://doi.org/10.1167/12.4.12

Estes, W. K. (1956). The problem of inference from curves based on group data. *Psychological Bulletin*, *53*(2), 134–140. https://doi.org/10.1037/h0045156

Geisler, W. S. (2011). Contributions of ideal observer theory to vision research. *Vision Research*, *51*(7), 771–781. https://doi.org/10.1016/j.visres.2010.09.027

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*(4), 457–472. https://doi.org/10.1214/ss/1177011136

Gibson, J. J. (1937). Adaptation, after-effect, and contrast in the perception of tilted lines. II. Simultaneous contrast and the areal restriction of the after-effect. *Journal of Experimental Psychology*, *20*(6), 553–569. https://doi.org/10.1037/h0057585

Gibson, J. J., & Radner, M. (1937). Adaptation, after-effect and contrast in the perception of tilted lines. I. Quantitative studies. *Journal of Experimental Psychology*, *20*(5), 453–467. https://doi.org/10.1037/h0059826

Golomb, J. D. (2015). Divided spatial attention and feature-mixing errors. *Attention, Perception & Psychophysics*, *77*(8), 2562–2569. https://doi.org/10.3758/s13414-015-0951-0

Golomb, J. D., L'heureux, Z. E., & Kanwisher, N. (2014). Feature-binding errors after eye movements and shifts of attention. *Psychological Science*, *25*(5), 1067–1078. https://doi.org/10.1177/0956797614522068

Guerin, S. A., Robbins, C. A., Gilmore, A. W., & Schacter, D. L. (2012). Retrieval failure contributes to gist-based false recognition. *Journal of Memory and Language*, *66*(1), 68–78. https://doi.org/10.1016/j.jml.2011.07.002

Haberman, J., & Whitney, D. (2009). Seeing the mean: Ensemble coding for sets of faces. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(3), 718–734. https://doi.org/10.1037/a0013899

Heathcote, A., Brown, S., & Mewhort, D. (2000). The power law repealed: The case for an exponential law of practice. *Psychonomic Bulletin & Review*, *7*(2), 185–207. https://doi.org/10.3758/bf03212979

Hemmer, P., & Steyvers, M. (2009). A Bayesian account of reconstructive memory. *Topics in Cognitive Science*, *1*(1), 189–202. https://doi.org/10.1111/j.1756-8765.2008.01010.x

Hiris, E., & Blake, R. (1996). Direction repulsion in motion transparency. *Visual Neuroscience*, *13*(1), 187–197. https://doi.org/10.1017/s0952523800007227

Honig, M., Ma, W. J., & Fougnie, D. (2020). Humans incorporate trial-to-trial working memory uncertainty into rewarded decisions. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(15), 8391–8397. https://doi.org/10.1073/pnas.1918143117

Huang, J., & Sekuler, R. (2010). Distortions in recall from visual memory: Two classes of attractors at work. *Journal of Vision*, *10*(10), 24. https://doi.org/10.1167/10.10.24

Huttenlocher, J., Hedges, L. V., & Duncan, S. (1991). Categories and particulars: Prototype effects in estimating spatial location. *Psychological Review*, *98*(3), 352–376. https://doi.org/10.1037/0033-295X.98.3.352

Huttenlocher, J., Hedges, L. V., & Vevea, J. L. (2000). Why do categories affect stimulus judgment? *Journal of Experimental Psychology: General*, *129*(2), 220–241. https://doi.org/10.1037/0096-3445.129.2.220

Jiang, Y., Olson, I. R., & Chun, M. M. (2000). Organization of visual short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(3), 683–702. https://doi.org/10.1037/0278-7393.26.3.683

Johnson, J. S., Simmering, V. R., & Buss, A. T. (2014). Beyond slots and resources: Grounding cognitive concepts in neural dynamics. *Attention, Perception & Psychophysics*, *76*(6), 1630–1654. https://doi.org/10.3758/s13414-013-0596-9

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*(5), 568–577. https://doi.org/10.1111/j.1467-9280.2009.02329.x

Konkle, T., & Oliva, A. (2012). A real-world size organization of object responses in occipitotemporal cortex. *Neuron*, *74*(6), 1114–1124. https://doi.org/10.1016/j.neuron.2012.04.036

Kruschke, J. K. (2014). *Doing Bayesian data analysis: A tutorial with r, jags, and stan*. Academic Press.

Lee, M. D., & Wagenmakers, E.-J. (2014). *Bayesian cognitive modeling: A practical course*. Cambridge University Press.

Levinson, E., & Sekuler, R. (1976). Adaptation alters perceived direction of motion. *Vision Research*, *16*(7), 779–IN7.

Lewandowsky, S., & Oberauer, K. (2009). No evidence for temporal decay in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(6), 1545–1551. https://doi.org/10.1037/a0017010

Lindley, D. V. (1965). *Introduction to probability and statistics from a Bayesian point of view, Pt. 2: Inference*. Cambridge University Press.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*(6657), 279–281. https://doi.org/10.1038/36846

Luck, S. J., & Vogel, E. K. (2013). Visual working memory capacity: From psychophysics and neurobiology to individual differences. *Trends in Cognitive Sciences*, *17*(8), 391–400. https://doi.org/10.1016/j.tics.2013.06.006

Marshak, W., & Sekuler, R. (1979). Mutual repulsion between moving visual targets. *Science*, *205*(4413), 1399–1401. https://doi.org/10.1126/science.472756

Mather, G. (1980). The movement aftereffect and a distribution-shift model for coding the direction of visual movement. *Perception*, *9*(4), 379–392. https://doi.org/10.1068/p090379

Navalpakkam, V., & Itti, L. (2007). Search goal tunes visual features optimally. *Neuron*, *53*(4), 605–617. https://doi.org/10.1016/j.neuron.2007.01.018

Newman, E. J., & Lindsay, D. S. (2009). False memories: What the hell are they for? *Applied Cognitive Psychology*, *23*(8), 1105–1121. https://doi.org/10.1002/acp.1613

O'Toole, B., & Wenderoth, P. (1977). The tilt illusion: Repulsion and attraction effects in the oblique meridian. *Vision Research*, *17*(3), 367–374. https://doi.org/10.1016/0042-6989(77)90025-6

Oberauer, K., & Kliegl, R. (2006). A formal model of capacity limits in working memory. *Journal of Memory and Language*, *55*(4), 601–626. https://doi.org/10.1016/j.jml.2006.08.009

Oberauer, K., & Lewandowsky, S. (2008). Forgetting in immediate serial recall: Decay, temporal distinctiveness, or interference? *Psychological Review*, *115*(3), 544–576. https://doi.org/10.1037/0033-295X.115.3.544

Oberauer, K., & Lin, H.-Y. (2017). An interference model of visual working memory. *Psychological Review*, *124*(1), 21–59. https://doi.org/10.1037/rev0000044

Oberauer, K., Stoneking, C., Wabersich, D., & Lin, H.-Y. (2017). Hierarchical Bayesian measurement models for continuous reproduction of visual features from working memory. *Journal of Vision*, *17*(5), 11. https://doi.org/10.1167/17.5.11

Ohio Supercomputer Center. (1987). http://osc.edu/ark:/19495/f5s1ph73

Pertzov, Y., Bays, P. M., Joseph, S., & Husain, M. (2013). Rapid forgetting prevented by retrospective attention cues. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(5), 1224–1231. https://doi.org/10.1037/a0030947

Plummer, M. (2003). "JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling." In K. Hornik, F. Leisch, & A. Zeileis (Eds.), *Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003)*. https://www.r-project.org/conferences/DSC-2003/Proceedings/Plummer.pdf

Rauber, H. J., & Treue, S. (1998). Reference repulsion when judging the direction of visual motion. *Perception*, *27*(4), 393–402. https://doi.org/10.1068/p270393

Robertson, L. C. (2003). Binding, spatial attention and perceptual awareness. *Nature Reviews Neuroscience*, *4*(2), 93–102. https://doi.org/10.1038/nrn1030

Schacter, D. L., Guerin, S. A., & St. Jacques, P. L. (2011). Memory distortion: An adaptive perspective. *Trends in Cognitive Sciences*, *15*(10), 467–474. https://doi.org/10.1016/j.tics.2011.08.004

Schutte, A. R., & Spencer, J. P. (2009). Tests of the dynamic field theory and the spatial precision hypothesis: Capturing a qualitative developmental transition in spatial working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(6), 1698–1725. https://doi.org/10.1037/a0015794

Scolari, M., & Serences, J. T. (2009). Adaptive allocation of attentional gain. *The Journal of Neuroscience* , *29*(38), 11933–11942. https://doi.org/10.1523/JNEUROSCI.5642-08.2009

Scotti, P. S., Hong, Y., Golomb, J. D., & Leber, A. B. (2021). Statistical learning as a reference point for memory distortions: Swap and shift errors. *Attention, Perception, & Psychophysics*. Advance online publication. https://doi.org/10.3758/s13414-020-02236-3

Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. Chapman and Hall.

Simmering, V. R., & Spencer, J. P. (2008). Generality with specificity: The dynamic field theory generalizes across tasks and time scales. *Developmental Science*, *11*(4), 541–555. https://doi.org/10.1111/j.1467-7687.2008.00700.x

Simmering, V. R., Spencer, J. P., & Schöner, G. (2006). Reference-related inhibition produces enhanced position discrimination and fast repulsion near axes of symmetry. *Perception & Psychophysics*, *63*, 1027–1046.

Spencer, J. P., Simmering, V. R., Schutte, A. R., & Schöner, G. (2007). What does theoretical neuroscience have to offer the study of behavioral development? Insights from a dynamic field theory of spatial cognition. In J. M. Plumert & J. P. Spencer (Eds.), *The emerging spatial mind* (pp. 320–361). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195189223.003.0014

Sutterer, D. W., Foster, J. J., Adam, K. C., Vogel, E. K., & Awh, E. (2019). Item-specific delay activity demonstrates concurrent storage of multiple active neural representations in working memory. *PLoS Biology*, *17*(4), e3000239. https://doi.org/10.1371/journal.pbio.3000239

Treisman, A., & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology*, *14*(1), 107–141. https://doi.org/10.1016/0010-0285(82)90006-8

Vehtari, A., Gabry, J., Yao, Y., & Gelman, A. (2019). loo: Efficient leave-one-out cross-validation and WAIC for Bayesian models (R package version 2.1.0.). https://CRAN.R-project.org/package=loo

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, *27*(5), 1413–1432. https://doi.org/10.1007/s11222-016-9696-4

Vergauwe, E., Barrouillet, P., & Camos, V. (2009). Visual and spatial working memory are not that dissociated after all: A time-based resource sharing account. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(4), 1012.

Vul, E., & Rich, A. N. (2010). Independent sampling of features enables conscious perception of bound objects. *Psychological Science*, *21*(8), 1168–1175. https://doi.org/10.1177/0956797610377341

Watanabe, S. (2010). Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, *11*, 3571–3594.

Wei, Z., Wang, X.-J., & Wang, D.-H. (2012). From distributed resources to limited slots in multiple-item working memory: A spiking network model with normalization. *The Journal of Neuroscience*, *32*(33), 11228–11240. https://doi.org/10.1523/JNEUROSCI.0735-12.2012

Wenderoth, P., & Johnstone, S. (1988). The different mechanisms of the direct and indirect tilt illusions. *Vision Research*, *28*(2), 301–312. https://doi.org/10.1016/0042-6989(88)90158-7

Wenderoth, P., & Wiese, M. (2008). Retinotopic encoding of the direction aftereffect. *Vision Research*, *48*(19), 1949–1954. https://doi.org/10.1016/j.visres.2008.06.013

Wiese, M., & Wenderoth, P. (2007). The different mechanisms of the motion direction illusion and aftereffect. *Vision Research*, *47*(14), 1963–1967. https://doi.org/10.1016/j.visres.2007.04.010

Wilson, H. R., & Cowan, J. D. (1972). Excitatory and inhibitory interactions in localized populations of model neurons. *Biophysical Journal*, *12*(1), 1–24. https://doi.org/10.1016/S0006-3495(72)86068-5

Yo, C., & Wilson, H. R. (1992). Perceived direction of moving two-dimensional patterns depends on duration, contrast and eccentricity. *Vision Research*, *32*(1), 135–147. https://doi.org/10.1016/0042-6989(92)90121-X

Yoo, A. H., Klyszejko, Z., Curtis, C. E., & Ma, W. J. (2018). Strategic allocation of working memory resource. *Scientific Reports*, *8*(1), 16162. https://doi.org/10.1038/s41598-018-34282-1

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, *453*(7192), 233–235. https://doi.org/10.1038/nature06860

Received May 9, 2020
Revision received December 2, 2020
Accepted January 25, 2021 ■