Memory Fidelity Reveals Qualitative Changes in Interactions Between Items in Visual Working Memory

Zachary Lively[1], Maria M. Robinson[2], and Aaron S. Benjamin[1]

[1]Department of Psychology, University of Illinois at Urbana-Champaign, and [2]Department of Psychology, University of California San Diego

Abstract

Memory for objects in a display sometimes reveals attraction—the objects are remembered as more similar to one another than they actually were—and sometimes reveals repulsion—the objects are remembered as more different from one another. The conditions that lead to these opposing memory biases are poorly understood; there is no theoretical framework that explains these contrasting dynamics. In three experiments (each $N = 30$ adults), we demonstrate that *memory fidelity* provides a unifying dimension that accommodates the existence of both types of visual working memory interactions. We show that either attraction or repulsion can arise simply as a function of manipulations of memory fidelity. We also demonstrate that subjective ratings of fidelity predict the presence of attraction or repulsion on a trial-by-trial basis. We discuss how these results bear on computational models of visual working memory and contextualize these results within the literature of attraction and repulsion effects in long-term memory and perception.

Visual working memory (VWM) is a fundamental memory system that allows us to simultaneously maintain multiple visual representations when processing new visual input and problem solving (Baddeley et al., 2011) or to briefly maintain visual representations retrieved from long-term memory (Fukuda & Woodman, 2017). These properties of VWM make it essential to most everyday functions (Luck & Vogel, 2013). Although there is mainstream consensus on the range of phenomena that VWM encompasses, the question of how memoranda are represented in VWM remains an active research domain.

A core assumption of prominent theories and computational models of VWM is that memory representations in VWM are maintained independently of one another. This simplifying assumption is implicit in mainstream discrete-slot (e.g., Cowan, 2001), pure-resource (e.g., Bays & Husain, 2008; Robinson et al., 2020; van den Berg et al., 2012; Wilken & Ma, 2004), mixture (Zhang & Luck, 2008), sample-size (Smith et al., 2016), interference (Oberauer, 2017), and neural (Bays, 2018) models of VWM. However, recent work suggests that when multiple items are held in VWM, memory representations often change when items in memory are similar to each other. In the current work, we focused on two qualitatively different types of memory biases— namely, *attraction* and *repulsion* of memory representations. Attraction describes when items in VWM become more similar to, or "attract," one another (Bae & Luck, 2019b; Brady & Alvarez, 2011; Dubé et al., 2014; Huttenlocher et al., 1991). Conversely, repulsion

Corresponding Author:
Zachary Lively, University of Illinois at Urbana-Champaign, Department of Psychology
Email: z.a.lively@gmail.com

describes when items in VWM become more different from, or "repulse," one another (e.g., Bae & Luck, 2017; Golomb, 2015; Suzuki & Cavanagh, 1997). The conditions that lead to these qualitatively different memory interactions between similar items are poorly understood; there is no theoretical framework that permits researchers to explain or make predictions regarding processes that give rise to these contrasting dynamics between VWM representations. Because many real-world scenarios require maintenance and manipulation of stimuli that overlap one another in feature space (e.g., faces), such a theory is needed for developing ecologically applicable models of VWM.

The current work provides a theoretical framework for understanding these contrasting interactions. Specifically, we propose that these qualitatively different memory interactions exist on a continuum of memory fidelity; when memory fidelity is high, representations of similar items are repulsed away from one another, and when memory fidelity is low, representations of similar items are attracted toward one another.

Attraction and Repulsion Effects in VWM

Evidence for attraction effects between items in VWM was reported by Brady and Alvarez (2011). In that study, participants had to briefly remember circles of different sizes and colors. Importantly, some circles in the memory set shared a color. These authors found evidence that participants' memory of size was systematically biased toward the average size of the circles of matching color. Brady and Alvarez proposed that VWM operates in a hierarchical fashion, in which individual items can be grouped on the basis of shared low-level features or on the basis of statistics about the ensembles into which they are categorized (Brady & Alvarez, 2011). When memory representations are noisy, information about the ensemble can be leveraged to optimally reconstruct aspects of forgotten individual items (Brady & Alvarez, 2011; Dubé et al., 2014; Hemmer & Steyvers, 2009).

Researchers have shown evidence that similar memory representations in VWM can also repulse away from one another (e.g., Bae & Luck, 2017; Golomb, 2015). For example, Bae and Luck instructed participants to briefly remember the orientation of two sequentially presented bars that varied in similarity. These authors found that the reported orientation for items that were similar (< 90° apart) were repulsed away from one another relative to pairs that had more dissimilar orientations.

The purpose of this article is to reconcile these conflicting results with a single unifying framework. We

Statement of Relevance

Visual working memory (VWM) is a fundamental memory system used to briefly maintain and manipulate visual information. VWM representations are often thought of as being independent of one another; however, this assumption has been challenged by recent findings that similar-appearing items can distort each other in qualitatively different ways. Yet why such memory biases arise remains poorly understood. This lack of understanding is significant because many real-world situations require VWM storage of similar-appearing items, such as faces or objects. In this research, we developed and found consistent support for a novel theory that can be used to explain and predict these memory biases across a range of conditions. We demonstrated that weaker memories result in similar items being remembered as more similar than they were, whereas stronger memories result in similar items being remembered as more dissimilar than they were. This work takes an important step toward building ecological theories of VWM.

assume that when memory representations are of poor fidelity, it is optimal to use group-level information (such as the average color of items) to maintain information about individual items (cf. Brady & Alvarez, 2011; Hemmer & Steyvers, 2009). In contrast, when memory representations are of high fidelity, similar representations are biased in a way that individuates items, making them more distinct from one another and thus less confusable.

In Experiment 1, we tested our theory by directly manipulating VWM fidelity (via memory load) and assessing whether we could replicate these two qualitatively different effects within a single experimental design. Experiment 2 was motivated by prominent models of VWM, according to which representations of individual items may vary in their resolution solely because of observer-induced variance in attention across items in the array. We also capitalized on evidence that memory-fidelity judgments can measure graded differences in the fidelity of VWM representations (van den Berg et al., 2017). Specifically, in Experiment 2, we presented participants with the same number of items in each display and collected participants' subjective reports of memory fidelity (i.e., confidence) on a trial-by-trial basis for randomly probed items. We found that the perceived memory fidelity of individual items was linked to both attraction effects and repulsion effects—items that were

perceived to be retained with low fidelity were attracted toward one another, whereas items that were perceived to be of high fidelity were repulsed away from one another.

The data from all experiments as well as the code used for the analyses are available on OSF (https://osf .io/92zvr/). For all experiments, we preregistered the design, the analysis plan, all of our predictions, and each of the comparisons that was relevant to our hypotheses on OSF (https://osf.io/92zvr/).[1] Each experiment had a within-subjects design.

Experiment 1

In Experiment 1, we tested the prediction that interactions between similar items in VWM qualitatively change with manipulations of memory load.

Method

Participants. Forty-six participants were recruited from the subject pool at the University of Illinois at Urbana-Champaign. We used data from the first 30 participants who did not fit our exclusion criterion. Note that we mainly based our sample size on those used in previous studies demonstrating interactions between similar items in VWM (Bae & Luck, 2017; Brady & Alvarez, 2011; Golomb, 2015). The cited studies found repulsion or attraction effects with sample sizes that ranged from 16 to 21 participants; here, to increase the power of this experiment, we collected a sample size of 30 adult participants. This sample size afforded 80% power to detect a medium effect size ($d_z = $ ~0.4).

We adopted standard exclusion criteria (e.g., Golomb, 2015) in which participants were excluded if their mean guess rate was greater than 50% (the guess rate was estimated using the standard mixture model). In doing so, we were not making the theoretical assumption that the standard mixture model is the best descriptive model of VWM. Instead, we were using the standard mixture model as a measurement model to quantify responses that deviate substantially from individuals' average responses, that is, responses that are best approximated by a uniform distribution. Ultimately, five participants were excluded on the basis of this criterion. All participants reported normal or corrected-to-normal vision. The experiment was approved by the university's institutional review board.

We collected data from 46 participants because we did not know in advance how many participants would meet our inclusion criteria; we used data from the first 30 participants who met our inclusion criteria, and we did not analyze the remaining data. This analysis plan is exactly the same as the one we laid out in our preregistration.

Apparatus and stimuli. Stimuli were presented against a black background on a 48 cm × 27 cm LCD monitor with a refresh rate of 68 Hz and a screen resolution of 1,920 × 1,080 pixels. Responses were made with a mouse. The experimental program was written in MATLAB (The MathWorks, Natick, MA), using Psychophysics Toolbox extensions (Version 3; Brainard, 1997; Kleiner et al., 2007).

Memory stimuli were colored squares that were 1° × 1° of visual angle. Each square was spatially grouped with one other square in the display, and grouped pairs were separated by 0.44° of visual angle. In the high-memory-load condition, the three pairs of squares formed either a right-side up triangle (i.e., there was one pair of colored squares in the bottom left, one pair in the bottom right, and one pair centered at the top of the screen) or an upside-down triangle. In the high-memory-load condition, pairs were 3.05° and 2.86° of visual angle away from each other on the horizontal and vertical axes, respectively. In the low-memory-load condition, a single pair of squares was presented in the center of the top or bottom third of the array. The memory display was designed such that participants did not have to make eye movements in order to foveate items in the memory array. Participants used a chin rest or head rest that was approximately 50 cm away from the screen. Participants reported on their memory for color by clicking a continuous color wheel, which was based on the Commission Internationale de l'Éclairage (CIE) L*a*b (L = 54, a = 21.5, b = 11.5; radius = 49) color space. This color wheel is publicly available online (Suchow et al., 2013; https://visionlab .github.io/MemToolbox/). On each trial, the wheel was randomly rotated on its axis.

Procedure. Participants adjusted their seat so that their eyes were level with the center of the screen. Participants were instructed to fixate on the center of the screen throughout the presentation of the memory array. Participants completed a practice session of 15 trials before beginning the experiment.

Figure 1 shows an example sequence of events. Each trial began with a blank screen (500 ms). Next, participants were briefly shown a central fixation cross (500 ms). After the fixation cross, the memory array was displayed for 250 ms. On high-memory-load trials (50% of trials), participants were shown six items (i.e., three pairs) in the memory array. On low-memory-load trials, participants were shown two items (i.e., one pair). Memory fidelity was manipulated via memory load because all mainstream models predict that memory load changes the fidelity of VWM representations, although they postulate different mechanisms for how they do so (e.g., van den Berg et al., 2012; Wilken & Ma, 2004; Zhang & Luck, 2008). Within each memory-load condition, the squares within each pair were either
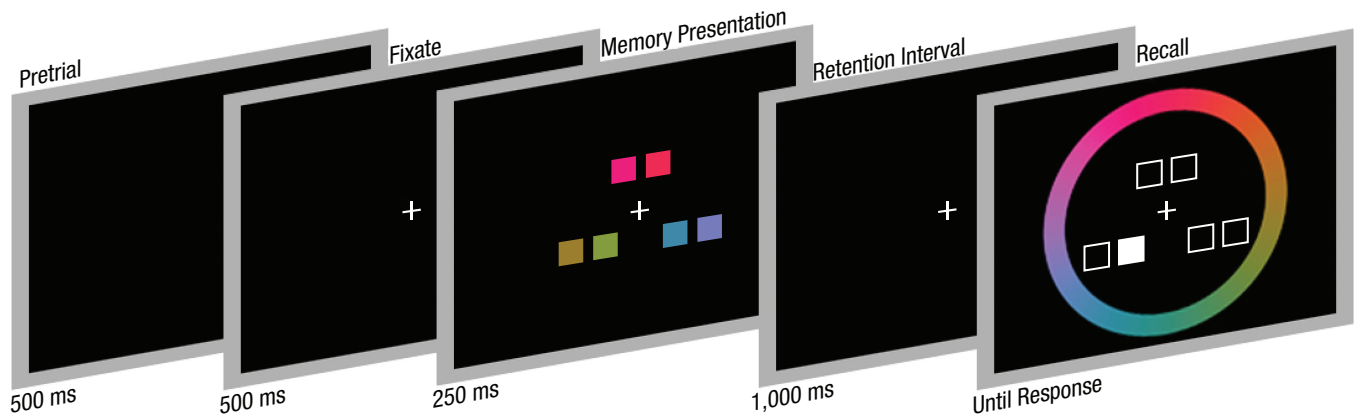
Fig. 1. Example trial sequence from Experiment 1. In the high-memory-load condition (shown here), participants were shown three pairs of to-be-remembered items (six items total). In the low-memory-load condition (not shown), participants were shown one pair of to-be-remembered items. Items within pairs could be either similar or dissimilar in terms of color space. After a retention interval, participants' task was to adjust a color wheel to match the color of the probed stimulus (indicated in the display by a white square).

similar (50% of trials) or dissimilar to each other in terms of color space. Pairs of items were presented in close spatial proximity to one another to enhance the effects of interest (Sahan et al., 2019). However, we note that repulsion and attraction can be obtained even when stimuli are not presented in close spatial proximity (e.g., Bae & Luck, 2017; Brady & Alvarez, 2011). Thus, the evidence does not suggest that the effects of interest hinge on similar items being presented close to one another in space. On trials in which paired items were similar to each other, they were 30° apart in CIE L*a*b color space, whereas on trials in which paired items were dissimilar to each other, they were 120° apart in color space.

The memory array was followed by a 1,000-ms retention interval. Next, participants were shown a color wheel and placeholders for all the items that were presented. All but one of the placeholders were white square frames (shown on a black background), and one placeholder was a solid white square, indicating its status as the probe stimulus. Participants were instructed to click on the part of the color wheel that matched the color of the square in the probed spatial location. Following standard procedure in VWM experiments, we randomly selected the color of the probed item on each trial from 360 colors, with the constraint that the same color was not probed two trials in a row. After every 104 trials, participants received a self-paced break that lasted a maximum of 60 s. There were 104 trials per condition, for a total of 416 trials. The order of conditions was randomized.

During the practice session, participants received feedback regarding how much their memory color deviated (in degrees) from the actual color of the probed item. During the experiment, no feedback was given.

Data analysis. Following previous work (e.g., Bae & Luck, 2017; Golomb, 2015), we measured memory bias by estimating the mean (μ) of the distribution of response errors from the memory test. We fitted three mainstream models of VWM to the data to assess whether our results generalized across different processing assumptions that one might make about VWM. Specifically, we used the *standard mixture model* (Zhang & Luck, 2008), the *swap model* (Bays et al., 2009), and a *simple resource model* (Green & Swets, 1966) to estimate μ and quantify participants' memory bias.

Memory responses were converted to difference scores (i.e., how far away the selected color was from the initially given color) in units of degrees. Because paired items could deviate from each other in the clockwise direction (50% of trials) or counterclockwise direction in color space, we took the additive inverse of all responses on trials in which the paired item deviated in the counterclockwise direction from the probed item. A mean difference score of zero indicated that there was no bias in participants' memory of the probed color. A positive difference score indicated that the participants' response tended (or was attracted) toward the item's spatially paired item, and a negative difference score indicated that a response tended (or was repulsed) away from the probed item's spatially paired item.

We fitted the three previously discussed VWM models to the distributions of difference scores separately for each participant and experimental condition using the *MemToolbox* (Suchow et al., 2013). According to the standard mixture model, the difference between the reported and actual color values of the probed item (i.e., error) is determined by the probability that an item is not in VWM and the participant has to randomly guess the probed item's color, combined with the

probability that the item is in VWM with some variable degree of resolution. Formally, this model is expressed by the following equation:

$$p\left(\theta; \gamma, \phi_{\mu,\kappa}\right) = \left(1 - \gamma\right) \times \phi_{\mu,\kappa} + \gamma \times \left(\frac{1}{2\pi}\right). \qquad (1)$$

Equation 1 shows the probability of reporting a value that deviates θ radians from the true color value, given parameters γ, μ, and κ, which denote the guess rate, the mean, and the concentration (analogous to variance in the normal distribution), respectively, of the von Mises distribution (denoted by ϕ). Note that the model is referred to as a mixture model because the probability density of errors on the continuous self-report task is given by a mixture of the von Mises and uniform distributions (the latter denoted by $\frac{1}{2\pi}$).

The swap model is a more complicated variant of the standard mixture model, in which people can confuse their memory of one item with the memory of another item in the memory array. This model assumes that participants' responses are determined by the probability that they remember a given item (and do not guess the item's color), the precision and bias with which they remember items when they are in memory, and the probability with which they swap items in memory. Formally, the swap model is given by the following equation:

$$p\left(\theta; \gamma, \phi_{\mu,\kappa}, \beta, \phi_{\mu_s,\kappa_s}\right) = \left(1 - \gamma - \beta\right) \times \phi_{\mu,\kappa} + \beta \times \phi_{\text{DistractorDistance},\kappa}$$
$$+ \gamma \times \left(\frac{1}{2\pi}\right). \qquad (2)$$

Note that Equation 2 is identical to Equation 1, with the exception that performance is also determined by the probability that participants swap items in VWM (denoted by β). *Distractor distance* refers to the distance of the distractors from the target item in color space.

Finally, we fitted our data with a VWM model that assumes that participants always have some information about the probed item and, accordingly, that participants' memory responses are never driven by guessing. This model fits within a class of resource models and is formally nothing more than a classic signal detection model (Green & Swets, 1966) applied to a circular space. The specific model we considered describes VWM as a resource that is distributed evenly across items and trials. Thus, according to this model, the distribution of response errors is determined by the precision with which a given item is coded as well as

by memory biases that can affect the central tendency of the distribution of errors. Formally, the probability density of errors on the continuous self-report task is represented with a von Mises distribution with parameters μ (measure of central tendency) and κ (measure of concentration), as shown in the following equation:

$$p\left(\theta; \phi_{\mu,\kappa}\right) = \phi_{\mu,\kappa}. \qquad (3)$$

Note that the circular mean is the maximum likelihood estimator of the mean of the von Mises distribution (Jammalamadaka & Sengupta, 2001).

Results

Manipulation of memory fidelity. We first evaluated whether our manipulation of memory load successfully affected memory fidelity. We used the concentration parameter (κ) to quantify memory fidelity across the three models; κ is analogous to the inverse of the variance of the von Mises distribution, and estimates of κ were converted to standard-deviation units. Larger standard deviations indicate more imprecise memories. Mean standard deviation was higher when memory load was high compared with when it was low, a result that was consistent across analyses using the mixture model, $F(1, 29) = 24.47$, $p < .001$, $\eta_p^2 = .46$, $\omega_p^2 = 0.43$; swap model, $F(1, 29) = 20.04$, $p < .001$, $\eta_p^2 = .41$, $\omega_p^2 = 0.38$; and resource model, $F(1, 29) = 20.04$, $p < .001$, $\eta_p^2 = .41$, $\omega_p^2 = .38$. This effect demonstrates that memory precision decreased with memory load, indicating that the manipulation of memory fidelity was successful. Additional analyses of the standard-deviation parameter are beyond the purposes of the present article and are reported in the Supplemental Material available online. Table 1 provides means, standard deviations, and 95% confidence intervals for the estimates of the standard-deviation parameter.

Memory bias (μ). The following analyses are planned comparisons (described in our preregistration) on μ, the measure of memory bias that is our primary dependent variable. Specifically, we assessed the effect of memory load and item similarity on memory bias, predicting that memory is biased toward similar but not dissimilar items when memory load is high, and memory is biased away from similar but not dissimilar items when memory load is low. We examined these effects with the three models described above. Ensuring that our conclusions survive varying plausible assumptions about processing dynamics indicates considerable robustness in those conclusions (Dutilh et al., 2019; Lee et al., 2019). We did not make predictions regarding how the other model parameters should change as a function of our combined

Lively et al.

Table 1. Descriptive Statistics for Estimates of Standard Deviation From Each Model and Condition in Experiment 1

Model and condition	*M*	*SD*	95% CI
Mixture			
Low: similar	20.77	3.82	[19.41, 22.14]
Low: dissimilar	18.13	3.72	[16.80, 19.46]
High: similar	33.46	13.76	[28.54, 38.38]
High: dissimilar	25.61	19.71	[18.56, 32.66]
Swap			
Low: similar	18.74	4.11	[17.27, 20.21]
Low: dissimilar	18.73	3.96	[17.31, 20.15]
High: similar	30.84	9.33	[27.50, 34.17]
High: dissimilar	24.47	15.69	[18.86, 30.09]
Resource			
Low: similar	25.68	8.18	[22.75, 28.61]
Low: dissimilar	31.32	9.51	[27.92, 34.73]
High: similar	69.96	14.27	[64.86, 75.07]
High: dissimilar	89.12	13.29	[84.27, 93.88]

Note: Low and high indicate memory-load conditions. Similar and dissimilar refer to conditions in which paired stimuli were similar or dissimilar to each other in color space. CI = confidence interval.

manipulations. Therefore, analyses of these dependent variables are exploratory. These additional analyses of the

standard deviation (transformation of κ), swap (γ), and guess rates (β) can be found in the Supplemental Material.

Figure 2 shows parameter estimates of μ. Table 2 summarizes means and standard deviations of μ and associated 95% confidence intervals. Figure 3 shows the raw distribution of response errors (in units of degrees) from all participants. The following *t* tests are planned comparisons described in our preregistration. As noted, we used μ estimates from each of the three models to quantify memory bias. Consistent with our predictions, results showed that in the low-memory-load condition, μ estimates were negative on average, indicating that similar items were repulsed away from each other in memory. In contrast, μ estimates in the high-memory-load condition were positive on average, indicating that similar items were attracted toward one another. These differences were obtained regardless of whether μ was estimated by the mixture model, $t(29) = 7.63$, $p < .001$, $d_z = 1.39$; swap model, $t(29) = 6.46$, $p < .001$, $d_z = 1.18$; or resource model, $t(29) = 5.65$, $p < .001$, $d_z = 1.03$.

There was no significant effect of memory load on memory bias when the probed item had a dissimilar partner. The results were also consistent across the three models—standard mixture model: $t(29) = 0.28$, $p = .78$,
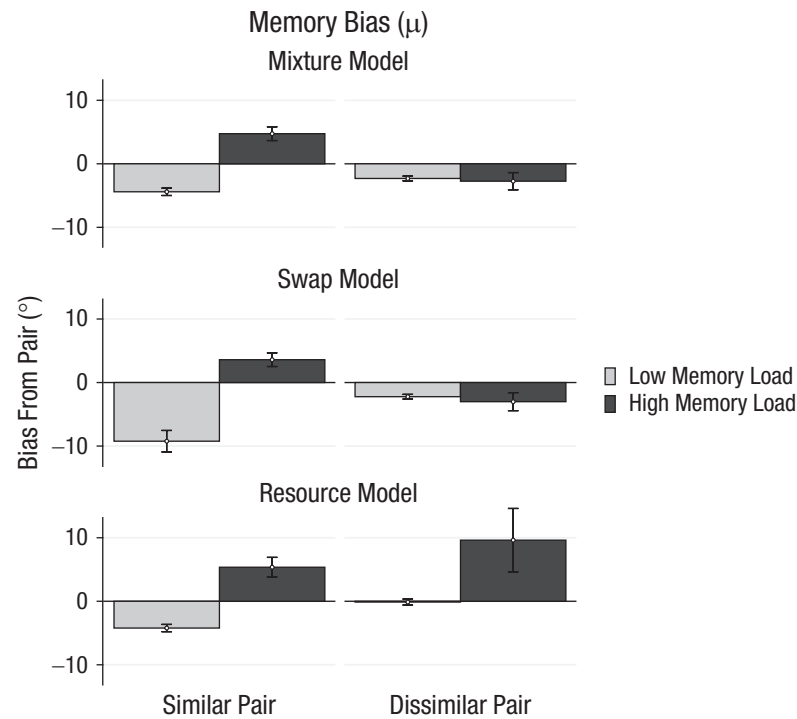

Fig. 2. Mean estimates of memory bias (μ) from Experiment 1. For each of the three models, memory bias is shown for similar and dissimilar items within pairs, separately for the high- and low-memory-load conditions. Positive values of μ indicate that the memory representation of the queried item was biased toward its paired item, and negative values of μ indicate that the memory representation was biased away from its paired item. Error bars show ±1 *SEM*.

$d_z = 0.05$, $BF_{01} = 6.82$; swap model: $t(29) = 0.51$, $p = .61$, $d_z = 0.09$, $BF_{01} = 6.24$; resource model: $t(29) = 1.97$, $p = .059$, $d_z = 0.36$, $BF_{01} = 1.2$. Note that in addition to our regular analyses, we also report Bayes factor (BF) estimates in favor of the null hypothesis (Rouder et al., 2009). BF estimates are reported using the Jeffrey-Zellner-Siow prior with scale on the effect size set to 1. Note that this statistic was not included in the preregistration but was proposed during the review process.

Finally, we found evidence supporting the prediction that repulsion was greater in the low-memory-load condition when paired items were similar than when they were dissimilar. We found these effects when μ was estimated by the mixture model, $t(29) = 2.99$, $p = .006$, $d_z = 0.55$; the swap model, $t(29) = 3.88$, $p < .001$, $d_z = 0.71$; and the resource model, $t(29) = 6.13$, $p < .001$, $d_z = 1.12$. In addition, participants showed greater attraction effects in the high-memory-load condition when paired items were similar compared with dissimilar, as estimated by the mixture model and swap model, $t(29) = 4.18$, $p < .001$, $d_z = 0.76$, and $t(29) = 3.82$, $p < .001$, $d_z = 0.7$, respectively. The generalization of this result to the swap model indicates that the effects of similarity on memory judgments cannot be explained simply by the idea that similar items were confused with one another. That is, we found these effects on memory bias even with a swap model that explicitly deconfounds such confusions from memory bias. Accounting for swap errors was especially important because these errors occur more frequently when items are in close spatial proximity to each other; likewise, swap errors can give the appearance of attraction effects (Matthey et al., 2015; Sahan et al., 2019). There was no difference in bias between the similar and dissimilar pairs in the high-memory-load condition when μ was estimated with the resource model, $t(29) = 0.83$, $p = .41$, $d_z = 0.15$, $BF_{01} = 5.08$.

Discussion

Experiment 1 provided support for the prediction that both attraction and repulsion in VWM representations can be induced via a manipulation of memory fidelity. Importantly, this pattern of results was consistent regardless of which model of VWM architecture we applied to estimate memory bias.

Experiment 2

Our aim in Experiment 2 was to generalize the results of Experiment 1 to a paradigm in which attraction and repulsion can be assessed within the same experimental condition and in which memory fidelity is determined on a subjective, trial-by-trial basis. This provides a strong test of generalizability because if self-reported

Table 2. Descriptive Statistics for Estimates of μ From Each Model and Condition in Experiment 1

Model and condition	M	SD	95% CI
Mixture			
Low: similar	−4.40	3.20	[−5.55, −3.26]
High: similar	4.74	5.94	[2.62, 6.86]
Low: dissimilar	−2.31	2.04	[−3.04, −1.58]
High: dissimilar	−2.74	7.41	[−5.39, −0.09]
Swap			
Low: similar	−9.24	9.27	[−12.55, −5.92]
High: similar	3.60	5.89	[1.49, 5.70]
Low: dissimilar	−2.23	2.01	[−2.95, −1.51]
High: dissimilar	−3.02	7.76	[−5.79, −0.24]
Resource			
Low: similar	−4.22	3.14	[−5.34, −3.09]
High: similar	5.36	8.50	[2.32, 8.40]
Low: dissimilar	−0.12	2.58	[−1.05, 0.80]
High: dissimilar	9.62	27.34	[−0.16, 19.41]

Note: Low and high indicate memory-load conditions. Similar and dissimilar refer to conditions in which paired stimuli were similar or dissimilar to each other in color space. CI = confidence interval.

memory fidelity tracks repulsion and attraction effects, we would expect those effects to be predictable either from manipulations that tax memory differentially (as in Experiment 1) or from subjective assessments of trial-to-trial variation in fidelity. In Experiment 2, we implemented only one level of memory load and collected participants' self-assessments of the accuracy with which they remembered the probed item on a trial-by-trial basis. We predicted attraction at low assessments of memory fidelity and repulsion at high assessments.

Experiment 2a

Method.

Participants. Forty-two participants were recruited from the subject pool at the University of Illinois at Urbana-Champaign. We used data from the first 30 participants who did not fit our exclusion criteria. As in Experiment 1, participants were excluded if their mean reported guessing rate was greater than 50% or if they did not complete the experiment. Additionally, participants were excluded if they contributed fewer than 25 trials to each memory-rating bin. This inclusion criterion was based on parameter-recovery simulations, which we used to determine the minimum number of trials necessary to recover unbiased estimates of μ, which is the dependent variable of interest. Critical in the current context is that parameter-recovery simulations are equivalent to a power analysis in the computational-modeling literature (Heathcote et al., 2015). For ease of exposition,
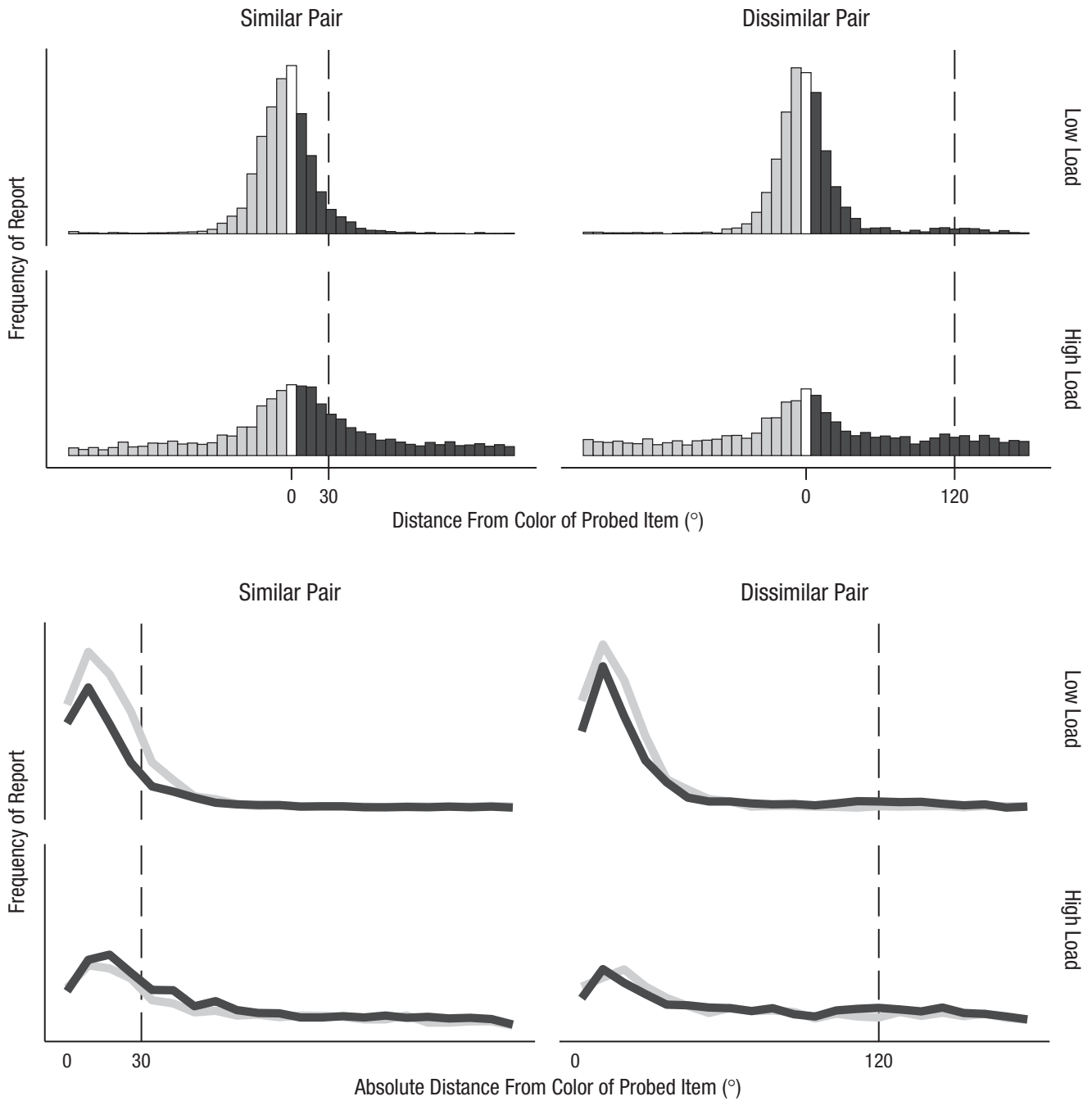
Fig. 3. Histograms (top) and frequency graphs (bottom) showing the distribution of raw responses from all participants in Experiment 1, separately for the low- and high-memory-load conditions. The dashed vertical line indicates the position of the paired item in color space (units of degrees) with respect to the queried item. The dark bars indicate responses that tend toward the paired item, and the light bars indicate responses that tend away from the paired item. The frequency graphs overlay the two (dark and light gray) halves of the histogram on the same section of the *x*-axis. In the absence of memory bias, the dark and light gray lines in the frequency plots would overlap completely. Instances in which the dark gray bar peaks above the light gray bar indicate attraction effects, whereas instances in which the light gray bar peaks above the dark gray bar indicate repulsion effects.

we provide detailed discussion of the parameter-recovery simulations in the Supplemental Material. We excluded 11 participants on the basis of these exclusion criteria. All participants reported normal or corrected-to-normal vision.

Procedure. To increase experimental power, we limited the design to the most informative similarity condition: item pairs that were similar to one another (i.e., 30° apart in color space). Thus, Experiment 2a used the same procedure as Experiment 1, except for the following

changes: (a) Participants were given a fixed memory load of six items on each trial, (b) participants were shown similar but not dissimilar pairs of items on every trial, and (c) after reporting on their memory via the color wheel, participants were asked to rate the fidelity of their memory for that item. Participants used the mouse to select one of three possible memory-fidelity ratings, which were as follows: (1) *guessing, or pretty close to just guessing*, (2) *response was of medium accuracy*, and (3) *response was highly accurate*.

Participants were given a 1-min break every 120 trials. There were 360 trials in total.

Data analysis. Memory responses were binned according to the memory rating they received. As before, we used the standard mixture, swap, and resource models to estimate memory bias (Bays et al., 2009; Green & Swets, 1966; Zhang & Luck, 2008).

For all ANOVAs, we used a Mauchly test for sphericity. In cases where sphericity was violated, we used a Greenhouse-Geisser correction and report a corrected *p* value (denoted *pc*).

Results.

Number of trials per memory-fidelity bin. Participants' ratings were spread symmetrically across the three memory-fidelity bins. On average, the medium-memory-fidelity bin contained the most trials ($M = 130.6$, $SD = 46.56$, minimum = 45, maximum = 231), the low-memory-fidelity bin contained the second most trials ($M = 115.5$, $SD = 59.4$, minimum = 27, maximum = 224), and the high-memory-fidelity bin contained the fewest trials ($M = 113.9$, $SD = 59.78$, minimum = 31, maximum = 288).

Manipulation of memory fidelity. A one-way repeated measures analysis of variance (ANOVA) revealed that there were significant differences in standard-deviation estimates across the three bins when standard deviation was estimated by the standard mixture model, $F(2, 58) = 7.45$, $pc = .008$, $\varepsilon = 0.57$, $\eta_p^2 = .20$, $\omega_p^2 = .17$; swap model, $F(2, 58) = 4.82$, $pc = .03$, $\varepsilon = 0.59$, $\eta_p^2 = .14$, $\omega_p^2 = .11$; and resource model, $F(2, 58) = 203.78$, $pc < .001$, $\eta_p^2 = .88$, $\omega_p^2 = .87$.

Follow-up *t* tests indicated that there was generally a significant difference in standard deviation across the three memory-fidelity bins; specifically, the standard deviation increased as memory fidelity decreased. Importantly, this result demonstrates that participants could track changes in their memory fidelity for individual items. There was a significant difference in standard deviation between high- and medium-memory-fidelity bins, when standard deviation was estimated by the standard mixture, swap, and resource models (all *p*s < .001). Likewise, there was a significant difference in standard deviation between high- and low-memory-fidelity bins when standard deviation was estimated by the standard

Table 3. Descriptive Statistics for Estimates of Standard Deviation From Each Model and Memory-Fidelity Bin in Experiment 2a

Model and memory-fidelity bin	*M*	*SD*	95% CI
Mixture			
Low	46.59	34.01	[34.42, 58.76]
Medium	39.14	15.81	[33.48, 44.8]
High	23.9	5.5	[21.94, 28.87]
Swap			
Low	37.11	26.59	[27.6, 46.63]
Medium	36.96	14.57	[31.75, 42.18]
High	23.75	5.27	[21.86, 25.63]
Resource			
Low	97.82	17.87	[91.43, 104.22]
Medium	61.06	22.81	[52.9, 69.22]
High	30.89	9.18	[27.61, 34.18]

Note: CI = confidence interval.

mixture model, $t(29) = 3.61$, $p = .001$, $d_z = 0.66$; swap model, $t(29) = 2.69$, $p = .01$, $d_z = 0.49$; and resource model, $t(29) = 21.67$, $p < .001$, $d_z = 3.96$. There was a significant difference between standard deviations in the low- and medium-memory-fidelity bins, when standard deviation was estimated by the resource model, $t(29) = 10.61$, $p < .001$, $d_z = 1.94$. There was no difference in standard deviation between these memory bins when standard deviation was estimated by the standard mixture model, $t(29) = 0.96$, $p = .35$, $d_z = 0.17$, $BF_{01} = 4.55$, or swap model, $t(29) = 0.02$, $p = .98$, $d_z = 0.004$, $BF_{01} = 7.08$. Table 3 provides means, standard deviations, and 95% confidence intervals for the estimates of the standard-deviation parameter.

Memory bias (μ). Figure 4 shows parameter estimates of μ. Table 4 summarizes means and standard deviations of μ and associated 95% confidence intervals. Figure 5 shows the raw distribution of response errors (in units of degrees) from all participants. As before, the following *t* tests are planned comparisons described in our preregistration, and we used μ estimates from each of the three models to quantify memory bias.

Memory-fidelity ratings did track changes in memory bias, but the outcomes only partially aligned with our predictions. Attraction effects were significantly larger in the medium-memory-fidelity bin than in the high-memory-fidelity bin when μ was estimated by the mixture model, $t(29) = 5.71$, $p < .001$, $d_z = 1.04$; swap model, $t(29) = 5.2$, $p < .001$, $d_z = 0.95$; and resource model, $t(29) = 4.79$, $p < .001$, $d_z = 0.88$. However, there was not a qualitative change in memory bias across these two bins (i.e., a switch from attraction to repulsion). There was no evidence of a difference between the low- and medium-memory-fidelity bins as estimated

by the mixture model, $t(29) = 1.22$, $p = .23$, $d_z = 0.22$, $\text{BF}_{01} = 3.49$; swap model, $t(29) = 2.02$, $p = .053$, $d_z = 0.37$, $\text{BF}_{01} = 1.1$; and resource model, $t(29) = 1.79$, $p = .08$, $d_z = 0.33$, $\text{BF}_{01} = 1.61$. There was also no difference between the low- and high-memory-fidelity bins when μ was estimated by the mixture model, $t(29) = 0.09$, $p = .93$, $d_z = 0.02$, $\text{BF}_{01} = 7.05$; swap model, $t(29) = 0.74$, $p = .46$, $d_z = 0.14$, $\text{BF}_{01} = 5.44$; and resource model, $t(29) = 0.86$, $p = .40$, $d_z = 0.16$, $\text{BF}_{01} = 4.96$. As described below, performance was very poor in the low-memory-fidelity bin; model estimates of guess rates were higher than 0.50 from the mixture and swap models, and estimates of standard deviation from the resource model were greater than 90° (estimates of model parameters other than μ are available in the Supplemental Material). It is likely that estimates of μ may have been too noisy in the low-memory-fidelity bin to yield stable estimates. Nevertheless, the difference between the medium- and high-memory-fidelity bins provides some evidence that perceived memory fidelity can be used to track the magnitude and presence of attraction effects.

Experiment 2b

Experiment 2a suggested that subjective memory fidelity can predict errors of attraction or repulsion.
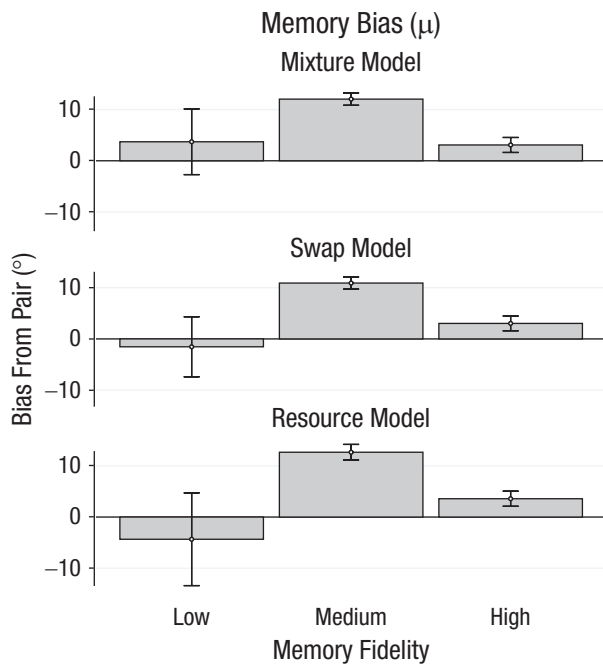

Fig. 4. Mean estimates of memory bias (μ) from Experiment 2a. For each of the three models, memory bias is shown for the three memory-fidelity bins. Positive values of μ indicate that the memory representation of the queried item was biased toward its paired item, and negative values of μ indicate that the memory representation was biased away from its paired item. Error bars show ± 1 *SEM*.

Table 4. Descriptive Statistics for Estimates of μ From Each Model and Memory-Fidelity Bin in Experiment 2a

Model and memory-fidelity bin	*M*	*SD*	95% CI
Mixture			
Low	3.71	35.14	[−8.86, 16.29]
Medium	12.04	6.57	[9.69, 14.39]
High	3.09	8.02	[0.22, 5.96]
Swap			
Low	−1.54	32.10	[−13.03, 9.95]
Medium	10.91	6.49	[8.59, 13.23]
High	3.03	7.87	[0.21, 5.84]
Resource			
Low	−4.34	49.59	[−22.08, 13.41]
Medium	12.66	8.38	[9.66, 15.66]
High	3.58	8.06	[0.69, 6.46]

Note: CI = confidence interval.

However, a full assessment of that prediction was hampered by the generally low performance across all confidence categories. To reevaluate the prediction under more auspicious conditions, we replicated this design with lower demands on memory in Experiment 2b.

Method. Thirty-five participants were recruited from the subject pool at the University of Illinois at Urbana-Champaign. We used the same exclusion criteria as in Experiment 2a. Three participants were excluded on the basis of these criteria.

The procedure of Experiment 2b was identical to that of Experiment 2a, with the exception that participants were presented with two pairs of similar items (i.e., four items total) instead of three. The data-analysis plan was also identical to the one used in Experiment 2a.

Results.

Number of trials per memory-fidelity bin. Experiment 2b followed the same criteria for participant exclusion as Experiment 2a, that is, each participant provided at least 25 trials per cell. On average, each participant had many more trials than this lower bound. Specifically, the medium-memory-fidelity bin had the most trials ($M = 150.2$, $SD = 54.27$, minimum = 36, maximum = 303), the high-memory-fidelity bin had the second most trials ($M = 132.83$, $SD = 54.95$, minimum = 27, maximum = 258), and the low-memory-fidelity bin had the fewest trials ($M = 77$, $SD = 36.6$, minimum = 25, maximum = 153).

Manipulation of memory fidelity. Participants' reported memory fidelity again tracked changes in the spread of their distribution of memory responses (*SD*). A one-way repeated measures ANOVA revealed that there were
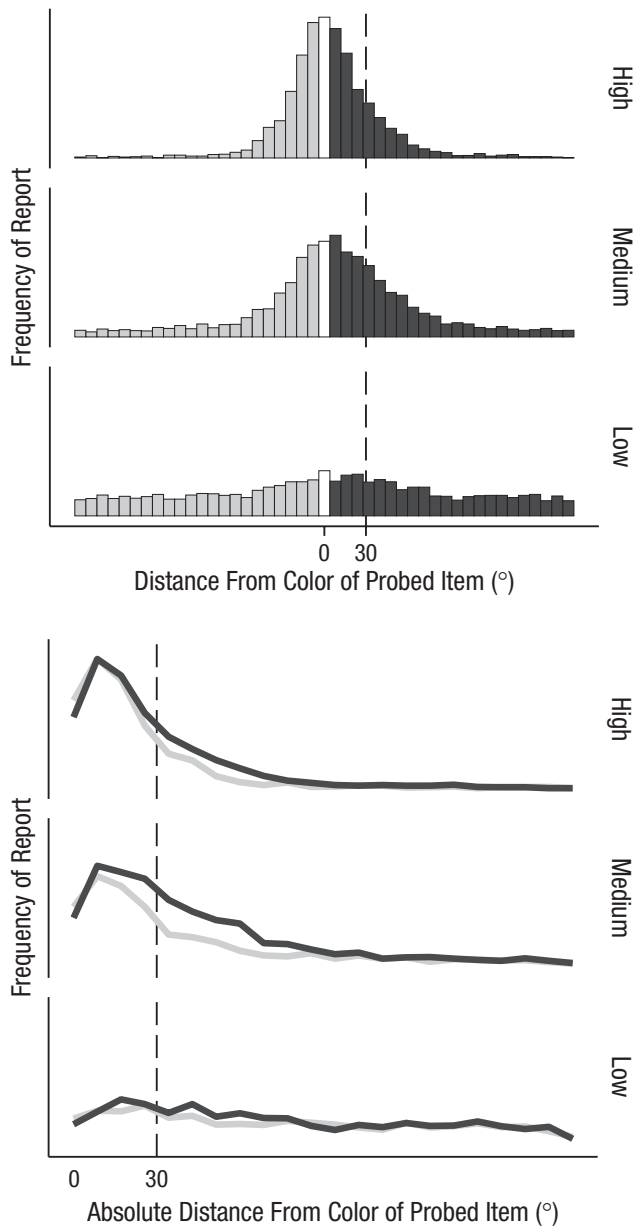
Fig. 5. Histograms (top) and frequency graphs (bottom) showing the distribution of raw responses from all participants in Experiment 2a, separately for each of the three memory-fidelity bins. The dashed vertical line indicates the position of the paired item in color space (units of degrees) with respect to the queried item. The dark bars indicate responses that tend toward the paired item, and the light bars indicate responses that tend away from the paired item. The frequency graphs overlay the two (dark and light gray) halves of the histogram on the same section of the x-axis. In the absence of memory bias, the dark and light gray lines in the frequency plots would overlap completely. Instances in which the dark gray bar peaks above the light gray bar indicate attraction effects, whereas instances in which the light gray bar peaks above the dark gray bar indicate repulsion effects.

significant differences in standard deviation across the three memory bins when standard deviation was estimated by the standard model, $F(2, 58) = 23.8$, $pc < .001$,

Table 5. Descriptive Statistics for Estimates of Standard Deviation From Each Model and Memory-Fidelity Bin in Experiment 2b

Model and memory-fidelity bin	M	SD	95% CI
Mixture			
Low	49.75	28.49	[39.55, 59.94]
Medium	28.96	5.65	[26.93, 30.98]
High	20.20	3.77	[18.85, 21.55]
Swap			
Low	51.92	26.88	[42.31, 61.54]
Medium	29.70	5.99	[27.55, 31.84]
High	20.68	3.82	[19.31, 22.05]
Resource			
Low	83.18	17.13	[77.04, 89.31]
Medium	46.18	13.75	[41.26, 51.1]
High	27.2	8.17	[24.28, 30.12]

Note: CI = confidence interval.

$\varepsilon = 0.52$, $\eta_p^2 = .45$, $\omega_p^2 = .43$; swap model, $F(2, 58) = 29.99$, $pc < .001$, $\varepsilon = 0.53$, $\eta_p^2 = .51$, $\omega_p^2 = .49$; and resource model, $F(2, 58) = 276.1$, $pc < .001$, $\varepsilon = 0.81$, $\eta_p^2 = .90$, $\omega_p^2 = .90$. Follow-up t tests indicated that standard deviations increased as perceived memory fidelity decreased and were significantly different between each memory bin for all three models (all $ps < .001$). These results, like the results of Experiment 2a, suggest that participants were able to accurately track and report changes in the fidelity of their memory representation. Table 5 provides means, standard deviations, and 95% confidence intervals for the estimates of the standard-deviation parameter.

Memory bias (μ). Figure 6 shows parameter estimates of μ. Table 6 summarizes means and standard deviations of μ and associated 95% confidence intervals. Figure 7 shows the raw distribution of response errors (in units of degrees) from all participants. As before, all comparisons of μ across memory-fidelity bins were planned. Consistent with our prediction, results showed that repulsion effects were evident when memory of items was rated as subjectively highly accurate, and attraction effects were evident at low levels of subjective memory fidelity. Specifically, we found a significant difference in μ between the high- and medium-memory-fidelity bins when μ was estimated via the mixture model, $t(29) = 7.81$, $p < .001$, $d_z = 1.43$; swap model, $t(29) = 4.54$, $p < .001$, $d_z = 0.81$; and resource model, $t(29) = 8.24$, $p < .001$, $d_z = 1.5$.

Critically, we also found evidence for *qualitative* changes in memory bias as a function of perceived memory fidelity. As shown in Table 6, μ estimates were significantly lower than zero when items were rated as being remembered with high accuracy. However, in the

medium- and low-memory-fidelity bins, the bias was in the direction of attraction. That is, average μ was significantly greater than zero when μ was estimated via the resource model, as indicated by the 95% confidence intervals. Average μ values were positive and quantitatively larger than zero when estimated from the mixture and swap models; however, estimates in the medium-memory-fidelity bin from these models were not significantly different from 0, as indicated by the 95% confidence intervals. Together, these results strongly support our prediction that memory biases can be tracked by an item's memory fidelity and may coexist within the same experimental condition.

There was a difference in μ between the high- and low-memory-fidelity bins when μ was estimated by the mixture model, $t(29) = 3.48$, $p = .002$, $d_z = 0.64$, and resource model, $t(29) = 2.7$, $p = .01$, $d_z = 0.49$. There was no difference in μ between the high- and low-memory-fidelity bins when μ was estimated by the swap model, $t(29) = 1.03$, $p = .31$, $d_z = 0.19$, $BF_{01} = 4.26$. There was also no difference in μ when we compared the low- and medium-memory-fidelity bins when μ was estimated by the standard mixture model, $t(29) = 1.02$, $p = .32$, $d_z = 0.19$, $BF_{01} = 4.3$; swap model, $t(29) = 0.46$, $p = .65$, $d_z = 0.08$, $BF_{01} = 6.39$; and resource model, $t(29) = 0.51$, $p = .61$, $d_z = 0.09$, $BF_{01} = 6.24$. Again, we point out that performance in the low-memory-fidelity

Table 6. Descriptive Statistics for Estimates of μ From Each Model and Memory-Fidelity Bin in Experiment 2b

Model and memory-fidelity bin	M	SD	95% CI
Mixture			
Low	4.37	15.14	[−1.04, 9.79]
Medium	1.55	5.76	[−0.51, 3.61]
High	−5.06	4.72	[−6.75, −3.37]
Swap			
Low	−0.90	15.82	[−6.56, 4.76]
Medium	0.53	5.94	[−1.6, 2.65]
High	−3.98	4.71	[−5.66, −2.29]
Resource			
Low	5.17	18.45	[−1.43, 11.77]
Medium	3.36	6.30	[1.11, 5.61]
High	−3.98	5.46	[−5.94, −2.03]

Note: CI = confidence interval.

bin was poor, which provides an explanation for why μ estimates in this memory bin are noisier within and across participants and do not yield stable estimates with or differences from other conditions. Nonetheless, all three models revealed that μ decreases with increasing memory fidelity, and two of the three models agreed that the path starts on the positive side and moves to the negative side.
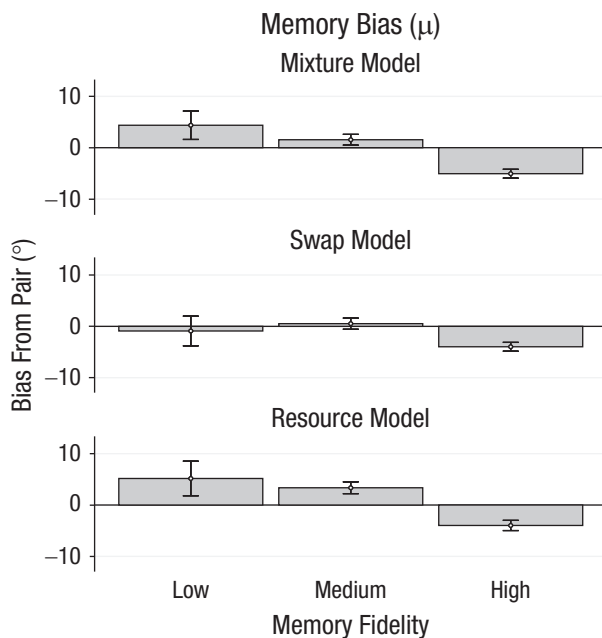
Discussion. In Experiment 2b, items perceived as being coded with high fidelity were repulsed away from similar items. In contrast, items coded with subjectively low fidelity attracted toward one another. The variation of memory fidelity across trials may reflect trial-by-trial variability of the distribution of an attention-like resource across items. Critically, Experiment 2 demonstrates that perceived memory fidelity tracks the presence and magnitude of attraction and repulsion effects even when actual display conditions are unchanging.

General Discussion

Results from Experiments 1 and 2 indicate that memory fidelity determines the nature of interactions between representations in VWM. When VWM representations are of poor fidelity, ensemble statistics convey the gist about low-level features of similar items and are depended on to avoid undue reliance on low-quality veridical representations of singular items. When VWM representation is of high fidelity, representations of individual items are biased to individuate memories that may otherwise interfere with one another because of their similarity. These qualitatively different interactions in VWM can arise from direct manipulations of memory

Fig. 6. Mean estimates of memory bias (μ) from Experiment 2b. For each of the three models, memory bias is shown for the three memory-fidelity bins. Positive values of μ indicate that the memory representation was biased toward its paired item, whereas negative values of μ indicate that the memory representation was biased away from its paired item. Error bars show ±1 *SEM*.

fidelity (Experiment 1) and even coexist within the same experimental condition (Experiment 2) because of natural fluctuations in processing, such as trial-by-trial variations in the allocation of attention to specific items (e.g., Patel et al., 2019).

The claim that attraction and repulsion effects in VWM are driven by memory fidelity situates these memory biases within a broader literature, including research in long-term memory and perception. As noted, ensemble representations and categorical information (Bae et al., 2015) can be used in the construction of more durable VWM representations (Brady & Alvarez, 2011). The same framework has been proposed in the long-term-memory literature. For instance, Hemmer and Steyvers (2009) proposed a Bayesian reconstructive model of long-term memory in which people optimally weight prior information about individual items. Gist information can be used to make weak memory representations more durable and more probabilistically accurate. We see these effects here as well.

These results also bear some similarity to serial-dependence attraction and repulsion effects in perception. Evidence suggests that the representation of a simple stimulus, such as a Gabor patch, may attract preceding representations (i.e., a positive serial-dependence effect). Notably, this effect is typically observed under conditions in which sensory signals are noisy (Cicchini et al., 2017) and representations are likely to be of low fidelity. New work also demonstrates evidence for repulsive serial dependence in VWM (Bae & Luck, 2019a, 2020). Repulsion effects are also apparent in serial perception, as evidenced by phenomena of aftereffects (e.g., Thompson & Burr, 2009) and perceptual adaptation; these effects appear to arise under conditions in which sensory signals are relatively strong (e.g., Gibson & Radner, 1937). There is no agreed-on framework for these qualitatively different serial-dependence effects, although several models have been proposed (Fritsche et al., 2020; Pascucci et al., 2019; Wei & Stocker, 2015). We assume the processes that give rise to interactions between VWM representations and interactions between serial percepts differ. However, behavioral data alone cannot directly speak to whether the biases we observe are strategic, perceptual, or some combination of both (see Yu & Geng, 2019). Relevant prior neural evidence is consistent with the idea that repulsion biases may reflect an optimal attention-induced bias in sensory-evoked responses in visual cortex (Scolari & Serences, 2009). Also, evidence suggests that behavioral measures of ensemble processing are correlated with increased activity in frontocentral areas of the brain, highlighting the potential role of
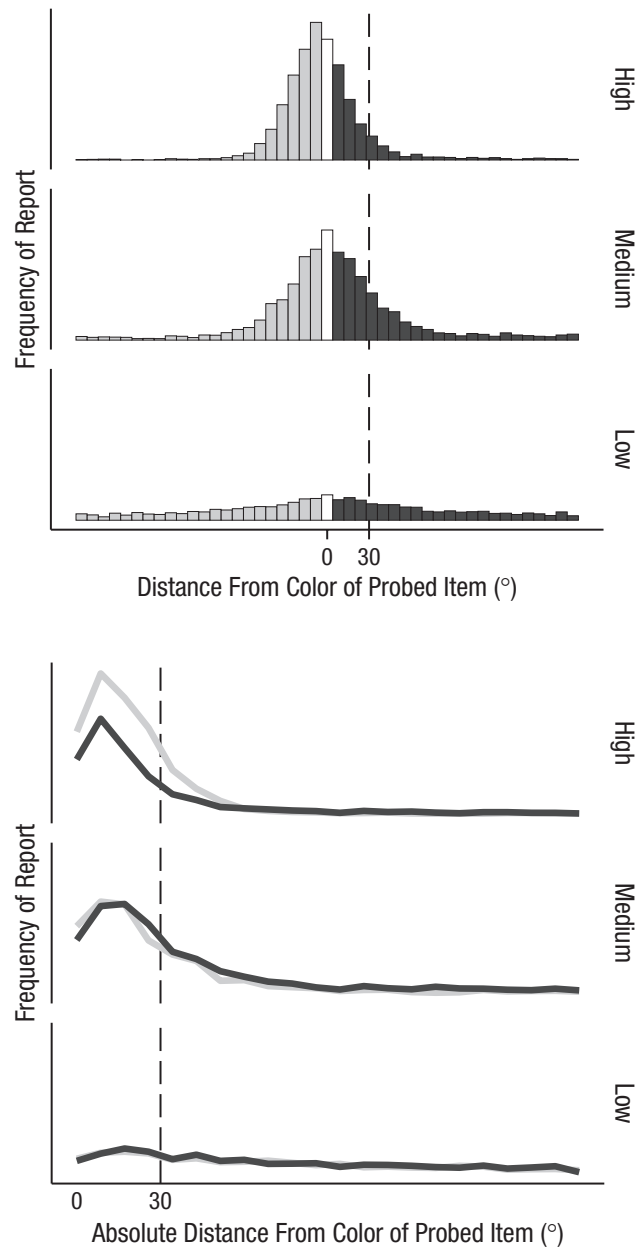

Fig. 7. Histograms (top) and frequency graphs (bottom) showing the distribution of raw responses from all participants in Experiment 2b, separately for each of the three memory-fidelity bins. The dashed vertical line indicates the position of the paired item in color space (units of degrees) with respect to the queried item. The dark bars indicate responses that tend toward the paired item, and the light bars indicate responses that tend away from the paired item. The frequency graphs overlay the two (dark and light gray) halves of the histogram on the same section of the *x*-axis. In the absence of memory bias, the dark and light gray lines in the frequency plots would overlap completely. Instances in which the dark gray bar peaks above the light gray bar indicate attraction effects, whereas instances in which the light gray bar peaks above the dark gray bar indicate repulsion effects.

abstract representations in ensemble processing (Oh et al., 2019). Future research may leverage neural measures to elucidate the mechanisms that give rise to these biases in VWM and how they vary with changes in memory fidelity.

In the course of conducting these experiments, it came to our attention that another group was conducting similar work in parallel (Chunharas et al., 2019). Relevant to the current framework, those authors found that repulsion effects correlated with participants' performance. Specifically, they found that participants who performed well on catch trials showed repulsion effects, whereas participants who performed poorly on catch trials did not show repulsion effects on the VWM task. That result corroborates the claim here that attraction and repulsion both result from processes related directly to memory fidelity.

Our results bear strong relevance to current efforts to model VWM architecture. To date, all prominent computational models of VWM share the tacit assumption that VWM representations are independent of one another. Our findings indicate that this assumption is incorrect, outside of the highly restrictive condition in which to-be-remembered objects are entirely dissimilar. Such conditions fail to capture the real-world deployment of VWM (Orhan & Jacobs, 2014), which is often based on extensive knowledge of statistical regularities in the natural environment (e.g., Fei-Fei et al., 2007). The current framework can be used as a springboard for future modeling work, which may capture quantitatively how prior information regarding interitem similarity is weighted to bias memory representations under varying VWM demands.

Finally, we note that our conclusion that memory fidelity gives rise to memory biases (rather than the reverse) is grounded on an extensive theoretical literature (e.g., Brady & Alvarez, 2011; Scolari & Serences, 2009). In the context of this literature, our interpretation provides a parsimonious explanation for why we observed distinct memory biases with variations in memory load, for instance. This view certainly does not rule out the idea that when memory biases arise, they have downstream consequences on the ongoing encoding of events that exacerbate or mitigate these tendencies.

Conclusion

We propose that the fidelity of representations in VWM determines how similar representations in VWM interact with one another. Ultimately, VWM is a limited-capacity system and thus inevitably has the potential for information loss. The documented memory biases may preemptively deal with potential loss of information as well as compensate for instances in which such loss is inevitable.

Transparency

Supplemental Material

Additional supporting information can be found at http://journals.sagepub.com/doi/suppl/10.1177/0956797621997367

Note

1. Although the preregistrations were all uploaded to OSF prior to data collection (as can be seen from the time stamps), they were not officially registered at that time; however, as the activity log shows, they have not been altered since being initially uploaded, and the project was registered prior to posting of the finished article.

References

Baddeley, A. D., Allen, R. J., & Hitch, G. J. (2011). Binding in visual working memory: The role of the episodic buffer. *Neuropsychologia*, *49*, 1393–1400.

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

Bae, G. Y., & Luck, S. J. (2019a). Reactivation of previous experiences in a working memory task. *Psychological Science*, *30*(4), 587–595.

Bae, G. Y., & Luck, S. J. (2019b). What happens to an individual visual working memory representation when it is interrupted? *British Journal of Psychology*, *110*(2), 268–287.

Bae, G. Y., & Luck, S. J. (2020). Serial dependence in vision: Merely encoding the previous-trial target is not enough. *Psychonomic Bulletin & Review*, *27*, 293–300.

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.

Bays, P. M. (2018). Reassessing the evidence of capacity limits in neural signals related to working memory. *Cerebral Cortex*, *28*, 1432–1438.

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), Article 7. https://doi.org/10.1167/9.10.7

Bays, P. M., & Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science*, *321*(5890), 851–854.

Brady, T. F., & Alvarez, G. A. (2011). Hierarchical encoding in visual working memory: Ensemble statistics bias memory for individual items. *Psychological Science*, *22*(3), 384–392.

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

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

Cicchini, G. M., Mikellidou, K., & Burr, D. (2017). Serial dependencies act directly on perception. *Journal of Vision*, *17*(14), Article 6. https://doi.org/10.1167/17.14.6

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*, 87–114.

Dubé, C., Zhou, F., Kahana, M. J., & Sekuler, R. (2014). Similarity-based distortion of visual short-term memory is due to perceptual averaging. *Vision Research*, *96*, 8–16.

Dutilh, G., Annis, J., Brown, S. D., Cassey, P., Evans, N. J., Grasman, R. P., & Kupitz, C. N. (2019). The quality of response time data inference: A blinded, collaborative assessment of the validity of cognitive models. *Psychonomic Bulletin & Review*, *26*(4), 1051–1069.

Fei-Fei, L., Iyer, A., Koch, C., & Perona, P. (2007). What do we perceive in a glance of a real-world scene? *Journal of Vision*, *7*(1), Article 10. https://doi.org/10.1167/7.1.10

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

Fukuda, K., & Woodman, G. F. (2017). Visual working memory buffers information retrieved from visual long-term memory. *Proceedings of the National Academy of Sciences, USA*, *114*(20), 5306–5311.

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*, 453–467.

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

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics* (Vol. 1). Wiley.

Heathcote, A., Brown, S. D., & Wagenmakers, E.-J. (2015). An introduction to good practices in cognitive modeling. In B. U. Forstmann & E.-J. Wagenmakers (Eds.), *An introduction to model-based cognitive neuroscience* (pp. 25–48). Springer.

Hemmer, P., & Steyvers, M. (2009). A Bayesian account of reconstructive memory. *Topics in Cognitive Science*, *1*(1), 189–202.

Huttenlocher, J., Hedges, L. V., & Duncan, S. (1991). Categories and particulars: Prototype effects in estimating spatial location. *Psychological Review*, *98*(3), 352–376.

Jammalamadaka, S. R., & Sengupta, A. (2001). *Topics in circular statistics* (Vol. 5). World Scientific.

Kleiner, M., Brainard, D. H., & Pelli, D. G. (2007). What's new in Psychtoolbox-3? *Perception*, *36*(ECVP Abstract Suppl.).

Lee, M. D., Criss, A. H., Devezer, B., Donkin, C., Etz, A., Leite, F. P., & Vandekerckhove, J. (2019). Robust modeling in cognitive science. *Computational Brain & Behavior*, *2*(3–4), 141–153.

Luck, S. J., & Vogel, E. K. (2013). Visual working memory capacity: From psychophysics and neurobiology to individual differences. *Trends in Cognitive Sciences*, *17*, 391–400.

Matthey, L., Bays, P. M., & Dayan, P. (2015). A probabilistic palimpsest model of visual short-term memory. *PLOS Computational Biology*, *11*(1), Article e1004003. https://doi.org/10.1371/journal.pcbi.1004003

Oberauer, K. (2017). An interference model of visual working memory. *Psychological Review*, *124*, 21–59.

Oh, B.-I., Kim, Y.-J., & Kang, M.-S. (2019). Ensemble representations reveal distinct neural coding of visual working memory. *Nature Communications*, *10*(1), Article 5665. https://doi.org/10.1038/s41467-019-13592-6

Orhan, A. E., & Jacobs, R. A. (2014). Toward ecologically realistic theories in visual short-term memory research. *Attention, Perception and Psychophysics*, *76*, 2158–2170. https://doi.org/10.3758/s13414-014-0649-8

Pascucci, D., Mancuso, G., Santandrea, E., Della Libera, C., Plomp, G., & Chelazzi, L. (2019). Laws of concatenated perception: Vision goes for novelty, decisions for perseverance. *PLOS Biology*, *17*(3), Article e3000144. https://doi.org/10.1371/journal.pbio.3000144

Patel, T. N., Steyvers, M., & Benjamin, A. S. (2019). Monitoring the ebb and flow of attention: Does controlling the onset of stimuli during encoding enhance memory? *Memory & Cognition*, *47*(4), 706–718.

Robinson, M. M., Benjamin, A. S., & Irwin, D. E. (2020). Is there a K in capacity? Assessing the structure of visual short-term memory. *Cognitive Psychology*, *121*, Article 101305. https://doi.org/10.1016/j.cogpsych.2020.101305

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian *t* tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*(2), 225–237.

Sahan, M. I., Dalmaijer, E. S., Verguts, T., Husain, M., & Fias, W. (2019). The graded fate of unattended stimulus representations in visuospatial working memory. *Frontiers in Psychology*, *10*, Article 374. https://doi.org/10.3389/fpsyg.2019.00374

Scolari, M., & Serences, J. T. (2009). Adaptive allocation of attentional gain. *The Journal of Neuroscience*, *29*(38), 11933–11942.

Smith, P. L., Lilburn, S. D., Corbett, E. A., Sewell, D. K., & Kyllingsbæk, S. (2016). The attention-weighted sample-size model of visual short-term memory: Attention capture predicts resource allocation and memory load. *Cognitive Psychology*, *89*, 71–105.

Suchow, J. W., Brady, T. F., Fougnie, D., & Alvarez, G. A. (2013). Modeling visual working memory with the MemToolbox. *Journal of Vision*, *13*(10), Article 9. https://doi.org/10.1167/13.10.9

Suzuki, S., & Cavanagh, P. (1997). Focused attention distorts visual space: An attentional repulsion effect. *Journal of Experimental Psychology: Human Perception and Performance*, *23*(2), 443–463.

Thompson, P., & Burr, D. (2009). Visual aftereffects. *Current Biology*, *19*, 11–14.

van den Berg, R., Shin, H., Chou, W. C., George, R., & Ma, W. J. (2012). Variability in encoding precision accounts for visual short-term memory limitations. *Proceedings of the National Academy of Sciences, USA*, *109*, 8780–8785.

van den Berg, R., Yoo, A. H., & Ma, W. J. (2017). Fechner's law in metacognition: A quantitative model of visual working memory confidence. *Psychological Review*, *124*, 197–214.

Wei, X.-X., & Stocker, A. A. (2015). A Bayesian observer model constrained by efficient coding can explain 'anti-Bayesian' percepts. *Nature Neuroscience*, *18*, 1509–1517.

Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision*, *4*(12), Article 11. https://doi.org/10.1167/4.12.11

Yu, X., & Geng, J. J. (2019). The attentional template is shifted and asymmetrically sharpened by distractor context. *Journal of Experimental Psychology: Human Perception and Performance*, *45*, 336–353.

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