AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: General

ISSN: 0096-3445

2024, Vol. 153, No. 1, 38–55
https://doi.org/10.1037/xge0001469

Judgments During Perceptual Comparisons Predict Distinct Forms of Memory Updating

Joseph M. Saito[1], Gi-Yeul Bae[2], and Keisuke Fukuda[1, 3]
[1] Department of Psychology, University of Toronto Mississauga
[2] Department of Psychology, Arizona State University
[3] Department of Psychology, University of Toronto Mississauga

Comparing a visual memory with new visual stimuli can bias memory content, especially when the new stimuli are perceived as similar. Perceptual comparisons of this kind may play a mechanistic role in memory updating and can explain how memories can become erroneous in daily life. To test this possibility, we investigated whether comparisons can produce other types of memory distortion beyond memory bias that are commonly implicated in erroneous memories (e.g., memory misattribution). We hypothesized that the type of memory distortion induced during a comparison depends on the perceived overlap between the memory and incoming stimulus—when the input is perceived as similar, it biases memory content; when perceived as the same, it replaces memory content. Participants completed a delayed estimation task in which they compared their memories of color (Experiment 1) and shape stimuli (Experiment 2) to probe stimuli before reporting memory content. We found systematic errors in participants' memory reports following perceived similarity and sameness that were toward the probes and larger following perceived sameness. Simulations confirmed that these errors were not explained by noisy encoding processes that occurred before comparisons. Instead, computational modeling suggested that these errors were likely explained by the probabilistic replacement of the memory by the probe following perceived sameness and integration between the memory and the probe following perceived similarity. Together, these findings suggest that perceptual comparisons can prompt distinct forms of memory updating that have been described previously and may explain how memories become erroneous during their use in everyday behavior.

Public Significance Statement
This study demonstrates that explicitly comparing one's memory of a visual object to a new object that is currently perceived risks distorting the memory representation. In particular, these findings show that if the observer judges the remembered object and new object to be similar to one another, the remembered object becomes more alike than the new object, and if they are judged to be identical, the new object replaces the remembered object in memory. Perceptual comparisons may therefore provide a mechanistic explanation for the formation of false memories in everyday life, including critical scenarios, such as eyewitness lineups.

Keywords: visual working memory, perceptual comparisons, memory bias, memory replacement

Supplemental materials: https://doi.org/10.1037/xge0001469.supp

Visual working memory (VWM) describes a collection of cognitive functions that allow for the temporary maintenance of prior visual inputs for accomplishing current tasks (Cowan, 2001). In particular, a great deal of research has illustrated the role of VWM in facilitating the recognition of a small amount of visual information that was previously seen (e.g., Luck & Vogel, 1997; Vogel

This article was published Online First August 31, 2023.

This research was supported by the Natural Sciences and Engineering Research Council (RGPIN-2017-06866) and the Connaught New Researcher Award. The authors have no conflicts of interest to declare that are relevant to the content of this article. The data are posted publicly at Open Science Framework (https://osf.io/d4q9h/). The data were previously disseminated as part of conference presentations at the 2021 Working Memory Symposium (https://www.wmsymposium.org/archive) and the 2021 Object Perception, Attention, & Memory (OPAM) Conference (https://www.opam.net/?page_id=30).

Joseph M. Saito served as lead for conceptualization, data curation, formal analysis, investigation, methodology, project administration, visualization, writing–original draft, and writing–review and editing. Gi-Yeul Bae served in a supporting role for formal analysis, methodology, supervision, and writing–review and editing. Keisuke Fukuda served as lead for funding acquisition and supervision and served in a supporting role for conceptualization, formal analysis, methodology, writing–original draft, and writing–review and editing.

Correspondence concerning this article should be addressed to Joseph M. Saito, Department of Psychology, University of Toronto Mississauga, 3359 Mississauga Road, Mississauga, ON L5L 1C6, Canada. Email: joseph.saito@mail.utoronto.ca

et al., 2001). This ability applies to perceptual inputs that are temporarily maintained in VWM and visual information that is recalled from long-term memory (LTM) back into VWM (e.g., Fukuda & Woodman, 2017; Sutterer et al., 2019; Vo et al., 2022). In this way, VWM serves as a temporary buffer that individuals use to perform perceptual comparisons between memory representations and current perceptual inputs.

However, recent findings have suggested that performing perceptual comparisons can invoke the updating of a VWM, even if doing so is not desired. Specifically, when asked to evaluate the perceptual similarity between a VWM item and a perceptual probe drawn from the same feature space, subsequent reports of the VWM item were systematically biased toward the perceptual probe, especially when the probe was perceived as similar (Fukuda et al., 2022). The researchers were able to show that these similarity-induced memory biases were explained computationally by representational integration between the memory and probe items that produced a reliable bias across trials. Later work then expanded upon these initial findings to show that biases induced during perceptual comparisons are distinct from those observed under other types of task demands. Specifically, when researchers directly compared report biases following perceptual comparisons to those observed when individuals perceived, but ignored novel inputs during VWM maintenance (e.g., Rademaker et al., 2015; Sun et al., 2017; Teng & Kravitz, 2019) and those observed when individuals maintained multiple memoranda in VWM (e.g., Chunharas et al., 2022; Scotti et al., 2021), biases were shown to be largest following perceptual comparisons, even after accounting for trial-wise differences in physical stimulus similarity and task differences in memory precision (Saito et al., 2022). Strikingly, this observed amplification of memory biases following perceptual comparisons was found to occur only on trials where the probe was perceived to be similar to the target, and not when it was perceived to be dissimilar, suggesting a causal role of perceived similarity in the bias modulation. The same researchers also found that memory biases following perceptual comparisons can persist across time in LTM, such that observers confidently report memory biases that are nearly identical in magnitude 24 hr after the comparison was performed (Saito et al., 2023). Together, this evidence has been used to posit that perceptual comparisons act as a cognitive mechanism that triggers memory updating and may explain systematic memory errors that are commonly observed in real-world scenarios, especially those where perceptual comparisons are performed explicitly as part of an ongoing task (e.g., eyewitness lineups; Steblay & Dysart, 2016; Wixted et al., 2016).

While these findings provide compelling evidence for a biasing effect of perceptual comparisons, some systematic memory errors are not amenable to a biased account. For example, many studies suggest that systematic memory errors can also arise when individuals misattribute novel perceptual details to a prior experience (Zaragoza & Lane, 1994). These source misattributions can lead to false memories of the prior experience in which novel details appear to replace those that were originally encoded (Brainerd & Reyna, 2005; Mitchell & Johnson, 2009) and are especially prominent when novel details are processed visually rather than verbally (e.g., Aizpurua et al., 2009; Braun & Loftus, 1998). If perceptual comparisons reflect a generalized cognitive mechanism that can explain a broad array of memory distortions, then comparing mnemonic and perceptual representations should also be capable of inducing memory replacement by perceptual inputs and not just memory bias.

But why might perceptual comparisons result in bias on some occasions and replacement on other occasions? As aforementioned, the perceived overlap between the memory and percept appears to play a direct role in determining the memory-updating effects that follow. In the case of perceived similarity and dissimilarity, the qualitative nature of memory updating is not changed between these judgments—both result in memory bias but with different magnitudes. Therefore, we reasoned that changing the qualitative nature of memory updating following comparison may require a change in psychological experience. Specifically, we predicted that perceiving sameness between memory and percept is psychologically different from perceiving similarity or dissimilarity and may induce a different type of memory updating that is akin to memory misattribution, namely, memory replacement.

When observers perceive a novel input to be similar or dissimilar to their current memory representation, this implies that the observer successfully detected differences between the representations and ascribed a unique identity to each of them. Conversely, when the observer perceives a novel input to be the same as their current memory representation, this implies that the observer failed to detect any differences between the representations and concluded that they were identical instead. Leading accounts of memory misattribution and misinformation effects suggest that failures to detect discrepancies between memories and novel details play a central role in making memories susceptible to these types of distortion (Butler & Loftus, 2018; Greene et al., 1982; Loftus, 1992; Thomas et al., 2010; Tousignant et al., 1986). Memory replacement following perceived sameness would also be in line with theoretical perspectives that suggest that memory distortions reflect unintended consequences of processing that is typically adaptive for behavior (e.g., Schacter et al., 2011). For example, in the case where a new input is in fact identical to one's memory, forming a fresh representation to replace the original memory may help prevent forgetting in the future.

To test this possibility, we conducted two experiments in which we asked participants to complete a delayed-estimation task in which they compared their VWM representation of a simple visual stimulus (i.e., color or shape) to a probe stimulus before reporting the VWM item. The probe stimulus was either identical to the encoded target or varied in its physical similarity. To preview our findings, we found that perceiving similarity or sameness in a different probe stimulus both resulted in systematic report errors towards the probe. Critically, these errors were considerably larger following incorrect "same" judgments than "similar" judgments—a behavioral pattern that would naturally arise if the probe was replacing the memory item, rather than integrating with it. We conducted simulations to address whether these systematic errors following "same" judgments were simply due to fluctuations in the quality of target encoding that were revealed by sorting trials based on participants' subjective judgments. We found that perceived sameness was dependent upon some minimum amount of representational overlap between the target and the probe, but "same" judgments were not merely identifying memories that were already alike the probes due to noisy encoding processes. We then performed a computational modeling analysis to offer a mechanistic explanation for the pattern of errors following "same" judgments. Specifically, we tested whether errors following "same" judgments were better explained by our hypothesized memory replacement mechanism or by the representational integration mechanism that has already been used to explain systematic errors following "similar" judgments (Fukuda et al.,

2022; Saito et al., 2022). Consistent with our predictions, we found unanimous evidence from our models that errors following "same" judgments were better explained by replacement than by integration while errors following "similar" judgments were better explained by integration than by replacement. Taken together, the present study provides convergent behavioral and computational evidence that perceptual comparisons can trigger qualitatively distinct types of memory updating depending on the perceived overlap between mnemonic and perceptual representations.

Experiments 1–2

To test the possibility that different judgments during perceptual comparisons invoke different memory-updating mechanisms, participants performed a delayed-estimation task in which they remembered a simple visual stimulus in VWM and compared it with a novel probe prior to a subsequent memory report. First, we hypothesized that VWM reports would be biased toward the novel probe following "similar" judgments, consistent with prior work suggesting that representational integration occurs following perceived similarity (Fukuda et al., 2022; Saito et al., 2022). More importantly, we hypothesized that systematic errors in the memory report would be larger following "same" judgments than "similar" judgments, consistent with the recruitment of a replacement mechanism instead. We conducted the experiment twice using a familiar (i.e., color, Experiment 1) and unfamiliar (i.e., shape, Experiment 2) type of visual stimulus, respectively, to ensure that our findings were not meaningfully contaminated by long-term categorical priors that have been shown to bias VWM representations (e.g., Bae et al., 2015).

Method

Transparency and Openness

We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study. All data and analysis code are available at the Open Science Framework (https://osf.io/d4q9h/). All behavioral and simulation analyses were conducted using MATLAB, Version R2020a (Mathworks, 2020) and the Psychophysics Toolbox extension, Version 3.0.16 (Kleiner et al., 2007). Computational modeling analyses were conducted using R, Version 4.0.2 (R Core Team, 2020). This study's design and analyses were not preregistered.

Participants

All participants in the experiment were undergraduate students at the University of Toronto Mississauga that reported normal or corrected-to-normal visual acuity and normal color vision. Each participant provided informed consent in accordance with the procedures approved by the Research Ethics Board at the University of Toronto.

We conducted a series of planned t tests to compare the magnitude of VWM errors following different judgments made during perceptual comparisons. Previous demonstrations of similarity-induced memory biases report large effect sizes (i.e., Cohen's $d > 0.8$; Fukuda et al., 2022; Saito et al., 2022). However, because we were primarily interested in investigating the potential for perceptual comparisons to induce memory replacement, which has not yet been investigated, we anticipated a more modest effect size (i.e., Cohen's $d = 0.6$). A power calculation performed with an alpha level of .05 and a statistical power of .9 indicated that we would need at least 32 subjects to obtain such an effect (Faul et al., 2007).

Participants were recruited on a weekly basis until the targeted number of subjects was reached. We recruited 49 participants in Experiment 1 (color) and 44 participants in Experiment 2 (shape). Each participant reported their demographic information via an online questionnaire in Qualtrics (Qualtrics Inc., 2020). For gender identity, participants selected between male, female, and prefer not to answer. For ethnic origin, participants reported the names of the countries that best describe their background. Racial identity was not collected. To extract reliable measures of memory precision and error, we assessed the proportion of memory reports made with high confidence in the baseline and experimental conditions for each participant. This led to the exclusion of five participants that failed to meet our a priori threshold of at least 15% confident trials in the baseline and experimental conditions (E1: four, E2: one) and one participant that did not report any memories confidently in either the baseline or experimental condition (E1: one). The remaining participants in the sample reported their memory with high confidence in more than 72% of trials in both the baseline and experimental conditions. Ten other participants were excluded for not following instructions (E2: two), failing to complete the perceptual comparison on at least 85% of experimental trials (E1: three, E2: one), poor overall task performance (memory precision in baseline condition $>3SD$ worse than the sample mean; E1: one, E2: two), and not finishing the experiment (E2: one). Of note, when we reconducted our analyses while including the individuals that reported a low number of confident trials or had poor overall task performance, all of the effects persisted (see the online supplemental materials). Data collected from the remaining 40 (31 female, nine male, $M_{age} = 19.7$ years old) and 37 participants (29 female, eight male, $M_{age} = 19.4$ years old) in Experiments 1 and 2, respectively, were submitted to analysis.

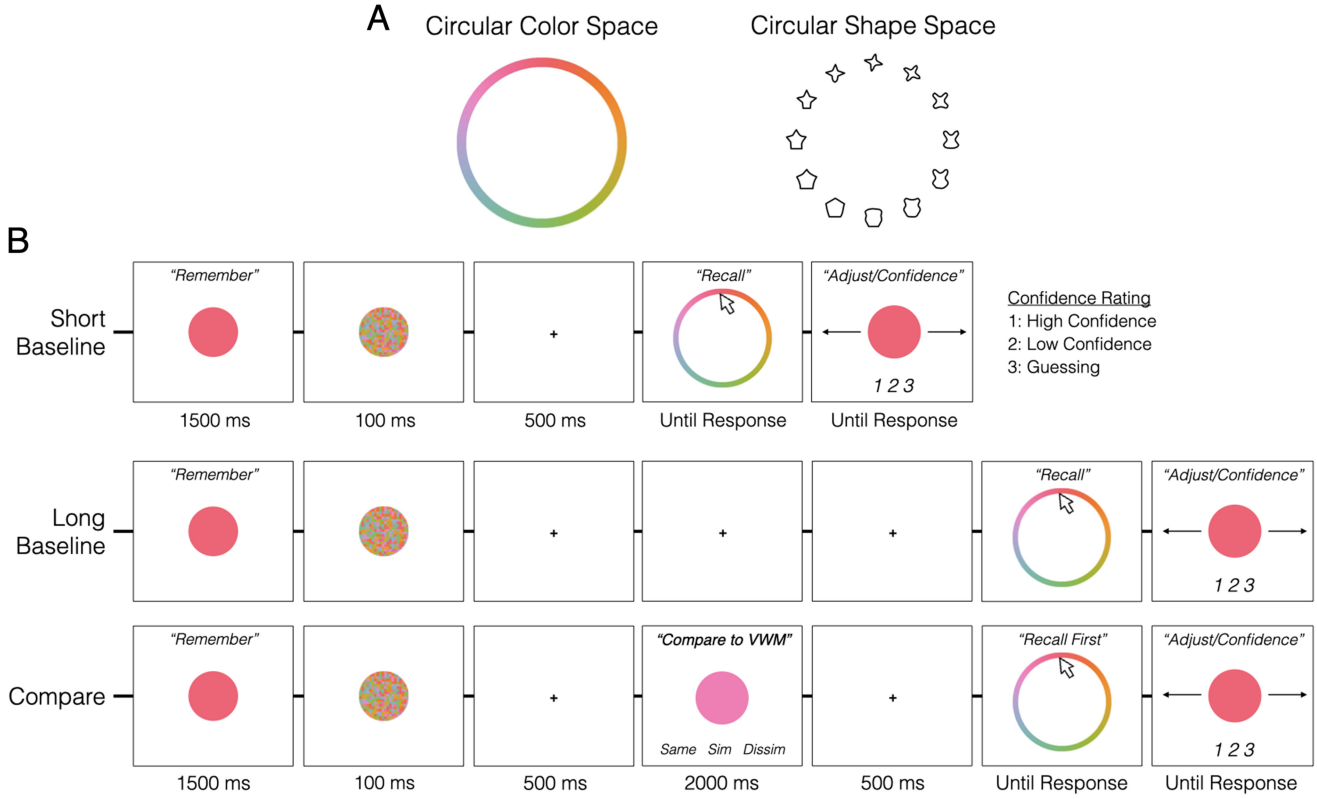
Apparatus and Stimuli

Participants completed the experiment remotely using a personal desktop or laptop computer. To ensure that experimental conditions were satisfactory, the completion of all task procedures was monitored by researchers in real-time using Zoom video conference software (Zoom Video Communications Inc., 2020). All stimuli were generated and presented in PsychoPy3 (Peirce, 2007), which was run locally on each participant's computer. Given that participants' viewing distance could not be tightly controlled using an online procedure, we report the fixed stimulus parameters in pixels and include the equivalent visual angle that would be assumed on a 1,920 × 1,080 pixel monitor with a 24-in. diagonal and a typical viewing distance of 60 cm.

For the color space, we sampled 360 equally-spaced color values from Commission Internationale de l´Eclairage $L^*a^*b^*$ space centered at $a^* = 20$ and $b^* = 38$ with a radius of 60. L^* was set to 70. The target and probe colors for a given trial were sampled from this color set and presented as circular color patches at 200 pixels (5.3°) in diameter. A circular color wheel was also created using this set of color values—such that each color value occupied 1° of the wheel—and was presented at 800 pixels (20.9°) in diameter (Figure 1A).

For the shape task, we used a continuous shape space whose circular visual similarity has been empirically validated (Figure 1A; Li

Figure 1
Experimental Schematic



Note. (A) Color and shape spaces. For illustration purposes, the shape space is shown with 12 exemplars. (B) The trial procedure for each baseline and experimental condition in the paradigm. See the online article for the color version of this figure.

et al., 2020). The shape space did not contain any prototypical shapes (e.g., triangles, squares) that could invoke long-term categorical priors. The memory and probe items for a given trial were sampled from 360 shapes within this stimulus set and presented at 200 × 200 pixels (5.3 × 5.3°). The shape wheel for a given trial consisted of 18 equidistant exemplar shapes presented in a circular arrangement at 720 pixels (19.9°) in diameter.

Procedure

For brevity, the description of the experimental task focuses on the use of color stimuli. All procedural differences between color and shape are explicitly noted.

Participants performed six blocks of 50 trials. Trials within each block were pseudorandomized between the baseline and experimental conditions (Figure 1B). Each trial began with a target color presented at the center of the screen for 1,500 ms, which participants were instructed to remember as precisely as possible. Target colors were randomly sampled from the circular color space. A visual mask was flashed for 100 ms immediately after the offset of the target color before the beginning of a maintenance interval that lasted 500 ms in the short baseline condition and 3,000 ms in the long baseline condition and the compare condition. At the completion of the maintenance interval, a circular color wheel was presented (Figure 1A). The color and shape wheels were randomly rotated on every trial in each experiment. Participants reported the original target color from their memory by moving the mouse to the part of the wheel where the target color was shown and clicking on the color. In the shape task, participants were told that the shape wheel consisted of 360 selectable shapes and that they could click in between the shapes displayed on the wheel if needed (see Apparatus and Stimuli section; Figure 1A). After selecting, a response probe was displayed at the center of the screen in the selected color. Participants were able to use the left and right arrow keys to fine-tune the color of the response probe to match what they remembered as precisely as possible. Afterward, they indicated their confidence in the accuracy of their memory report by pressing one of three keyboard buttons (high confidence, low confidence, guessing). The accuracy of the memory report was emphasized and was therefore reported without an imposed time limit.

In each trial of the *compare condition*, participants completed a perceptual comparison during the 3,000-ms maintenance interval (Figure 1B). Five hundred miliseconds after the offset of the mask, a novel probe color was presented at the center of the screen for 2,000 ms. The novel probe color was sampled ±0, 15, or 45° away from the memory item in the circular color space. Participants were instructed to compare the novel probe color to the target color being retained and judge whether the novel probe color was the same, similar, or dissimilar to the target color. Participants reported their judgment by pressing a corresponding button on the keyboard while the

probe color was onscreen (1 *for same*, 2 *for similar*, and 3 *for dissimilar*). The probe color remained on the screen for 2,000 ms regardless of the report and was followed by a 500-ms blank delay before participants reported the target color from memory.

The *short* and *long baseline conditions* each occurred in 20% of trials, respectively, and the *compare condition* occurred in the remaining 60%. Within the *compare condition*, the physical distance between the target and probe colors was counterbalanced across the three possible values indicated above (0, 15, 45°). The direction of offset from the target color was randomly assigned trial-by-trial.

Analyses

For every trial, we computed the response error during the memory report by subtracting the degree value of the memory report in the circular feature space from that of the original memory item. To provide directional information about the response error relative to the novel probe stimulus, we aligned the response errors across trials such that positive error values indicated response errors in the direction of the novel probe item (signed response errors). The direction of response errors in the baseline and identical probe conditions were randomly assigned. To quantify the size of the average memory error for each participant, we computed the mean signed response error for each condition. Memory precision was calculated by computing the inverse standard deviation (i.e., 1/*SD*) of the raw response errors within the relevant condition.

To maximize power in our analyses, we report results across all trials. However, when we repeated our analyses while including only trials where participants reported being highly confident in their memory report, all findings were preserved (see the online supplemental materials). We also report all statistical results with outliers included. Separate analyses confirmed that removing these outliers does not change any of the reported differences between conditions.

We report both frequentist (i.e., *t* values) and Bayesian (i.e., Bayes factors) statistics to allow for providing evidence in favor of null differences between conditions. BF_{01} indicates evidence in favor of the null hypothesis and BF_{10} indicates evidence in favor of the alternative hypothesis.

The data for this study are available at the Open Science Framework (https://osf.io/d4q9h/). This study was not preregistered.

Results

We began by separating trials according to the physical distance between the memory and the probe (0, 15, 45°) and the judgments made during perceptual comparisons ("same," "similar," "dissimilar"). As can be seen in Figure 2, there was correspondence between the physical distance separating the memory and probe and the perceived overlap reported by participants.

To begin investigating the presence of similarity-induced memory biases and the potential presence of an alternative replacement mechanism, we identified the physical distance bins that provided a reasonable proportion of trials for each respective judgment of interest. For example, we noted that the 45° distance bin contained a fair number of both "similar" and "dissimilar" judgments ("similar" color: $M = 18.48$ trials; "dissimilar" color: $M = 39.63$ trials, "similar" shape, $M = 22.19$ trials; "dissimilar" shape, $M = 33.14$ trials), thereby allowing us to test the replication of similarity-induced memory biases in the present data set while controlling for the physical distance between stimuli. To assess behavioral evidence consistent with our hypothesized replacement mechanism, we focused on the 15° distance bin which contained a fair number of "same" and "similar" judgments ("same" color, $M = 16.18$ trials; "similar" color, $M = 35.95$ trials, "same" shape, $M = 19.59$ trials; "similar" shape, $M = 34.27$ trials).

Replicating Similarity-Induced Memory Biases

To validate the efficacy of our perceptual comparison paradigm, we first sought to replicate the presence of similarity-induced memory biases in our data set. To test this, we compared the size of memory errors following "similar" judgments to those following "dissimilar" judgments in the 45° distance bin. If similarity-induced memory biases did occur, we should expect that errors following "similar" judgments were reliably larger than those following "dissimilar" judgments (Fukuda et al., 2022; Saito et al., 2022). Consistent with prior studies, we found evidence of reliable memory biases following "similar" judgments (Figure 3; color, $M = 10.02°$, 95% CI [7.26°, 12.78°], $t(38) = 7.36$, $p < .001$, Cohen's $d = 1.18$, $BF_{10} = 1.56 \times 10^6$; shape, $M = 12.19°$, [9.69°, 14.69°], $t(35) = 9.88$, $p < .001$, Cohen's $d = 1.65$, $BF_{10} = 7.94 \times 10^8$) and following "dissimilar" judgments in both stimulus types (Figure 3; color, $M = 3.40°$, [2.13°, 4.68°], $t(39) = 5.39$, $p < .001$, Cohen's $d = 0.85$, $BF_{10} = 5.07 \times 10^3$; shape, $M = 2.65°$, [0.18°, 5.11°], $t(36) = 2.18$, $p = .036$, Cohen's $d = 0.36$, $BF_{10} = 1.43$). Critically, biases following "similar" judgments were larger than those following "dissimilar" judgments (color, $M = 6.52°$, 95% CI [3.66°, 9.38°], $t(38) = 4.62$, $p < .001$, Cohen's $d = 0.74$, $BF_{10} = 5.25 \times 10^2$; shape, $M = 9.58°$, [6.63°, 12.53°], $t(35) = 6.60$, $p < .001$, Cohen's $d = 1.10$, $BF_{10} = 1.20 \times 10^5$).

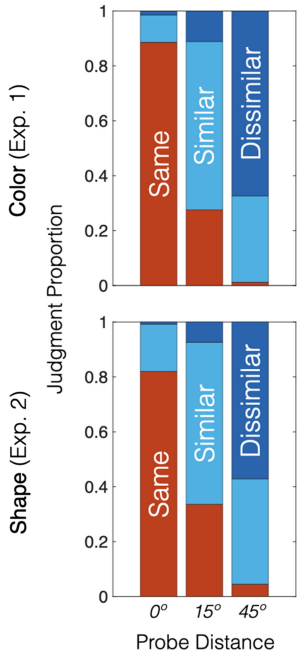
"Same" Judgments Result in Larger Memory Errors Than "Similar" Judgments

We then moved to assess whether our behavioral results showed patterns of memory errors following "same" judgments that were consistent with our hypothesized replacement mechanism. Specifically, if memory replacement occurred following "same" judgments, we should expect that errors following "same" judgments were larger than those following "similar" judgments. We found evidence of memory biases in the 15° distance bin following "similar" responses that reached significance for shape, but not for color (Figure 4; color, $M = 0.73°$, 95% CI [−0.46°, 1.92°], $t(39) = 1.24$, $p = .221$, Cohen's $d = 0.20$, $BF_{01} = 2.87$; shape, $M = 2.59°$, [1.45°, 3.73°], $t(36) = 4.62$, $p < .001$, Cohen's $d = 0.76$, $BF_{10} = 4.84 \times 10^2$).[1] We also found memory errors following "same" judgments that were reliable in both stimulus types (Figure 4; color, $M = 11.16°$, 95% CI [10.10°, 12.22°], $t(38) = 21.33$, $p < .001$, Cohen's $d = 3.42$, $BF_{10} = 3.89 \times 10^{19}$; shape, $M = 7.86°$, [6.32°, 9.40°], $t(36) = 10.34$, $p < .001$, Cohen's $d = 1.70$, $BF_{10} = 3.41 \times 10^9$). Consistent with our hypothesis, we found that the errors following "same" judgments were reliably larger than those following "similar" judgments (color, $M = 10.37°$, 95% CI [8.81°, 11.94°], $t(38) = 13.41$, $p < .001$, Cohen's $d = 2.15$,

[1] Given the sizable bias observed following "similar" judgments in the 45° distance bin (Figure 3), we reasoned that the bias in the 15° distance bin may have failed to reach significance for color stimuli because of the high physical proximity between the target and probe that naturally limited the magnitude of the bias following integration.

Figure 2

Judgment Proportions Across Physical Distances



Color (Exp. 1)	Subjective Judgment \ Probe Distance		0°	15°	45°
Dissimilar	Dissimilar	%	1.50 (6.01)	11.12 (12.70)	67.34 (23.65)
		n	0.88 (3.49)	6.50 (7.43)	39.63 (14.00)
Similar	Similar	%	9.92 (8.26)	61.31 (18.31)	31.47 (23.43)
		n	5.88 (4.89)	35.95 (10.80)	18.48 (13.73)
Same	Same	%	88.58 (11.37)	27.56 (15.00)	1.19 (1.55)
		n	52.50 (6.85)	16.18 (8.80)	0.70 (0.91)

Shape (Exp. 2)	Subjective Judgment \ Probe Distance		0°	15°	45°
Dissimilar	Dissimilar	%	0.80 (1.66)	7.38 (9.12)	57.11 (25.43)
		n	0.46 (0.96)	4.30 (5.33)	33.14 (15.12)
Similar	Similar	%	17.20 (12.46)	59.00 (18.29)	38.37 (25.67)
		n	10.00 (7.15)	34.27 (10.50)	22.19 (14.96)
Same	Same	%	82.00 (13.11)	33.62 (16.31)	4.52 (10.98)
		n	47.86 (7.85)	19.59 (9.58)	2.62 (6.55)

Note. (A) Stacked bar charts showing the mean proportion of "same," "similar," and "dissimilar" judgments across participants at each probe distance. Proportions indicated by the height of the bars are reported in (B) as mean percentages of trials within a distance bin (%) and mean trial counts (*n*) with respective standard deviations in parentheses. See the online article for the color version of this figure.

$BF_{10} = 1.12 \times 10^{13}$; shape, $M = 5.27°$, [3.30°, 7.25°], $t(36) = 5.41$, $p < .001$, Cohen's $d = 0.89$, $BF_{10} = 4.51 \times 10^{3}$).

Discussion

We predicted and found that "similar" and "same" judgments made during perceptual comparisons resulted in systematic memory errors in the direction of the probe. We found that systematic errors following "similar" judgments were larger than those following "dissimilar" judgments, confirming the presence of similarity-induced memory biases. More importantly, we found that errors following "same" judgments were larger than those following "similar" judgments, consistent with the behavioral patterns that would be expected if perceived sameness triggered memory replacement rather than memory-probe integration.

However, these behavioral results alone only provide tentative support for the existence of distinct memory-updating mechanisms. In addition to the present hypothesis, the observed difference in the size of memory errors between "same" and "similar" judgments may also be explained by a more trivial account in which memories were not changed at all following perceived sameness. Instead, "same" judgments may have merely identified memories that were noisily encoded to be like the probe before the perceptual comparison took place. In this case, sorting trials based on subjective judgments tracked differences in the quality of initial encoding rather than a distortion caused by the perceptual comparison (see Fukuda et al., 2022 for a direct investigation of this issue in "similar" judgments). In order to conclude that perceived sameness actually triggered memory updating, we first needed to rule out this encoding accuracy account.

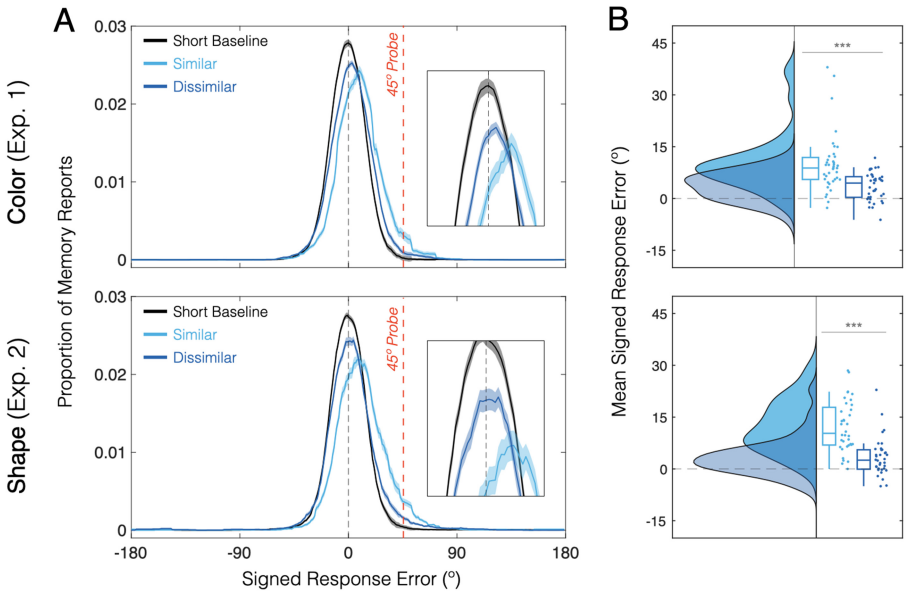
Simulating VWM Errors as Fluctuations in Encoding Accuracy

In order for a VWM and probe stimulus to be perceived as the same, the encoded target and probe representations must overlap with one another. If there is not a sufficient amount of overlap between the representations, it is unlikely that the target and probe will be perceived as identical and individuals will likely judge them to be "similar" or "dissimilar" instead. This prerequisite of representational overlap may be sufficient to explain the patterns of memory errors that we observed in behavior. That is, perceived sameness between memory and percept may not have triggered memory replacement, but instead identified instances where participants' VWM representation happened to be noisily encoded to be like the probe before the probe was even perceived. If this is the case, then the difference in memory errors observed following "same" and "similar" judgments does not reflect dissociable types of memory updating that were induced by perceptual comparisons, but instead reflects artificial systematicity in participants' memory reports that was introduced by sorting trials based on participants' subjective judgments (see Fukuda et al., 2022 for a direct investigation of this issue in "similar" judgments).

To rule out this explanation, we conducted a two-pronged analysis of participants' memory reports following "same" judgments. First, we show that memory replacement and the *encoding accuracy account* described above are not mutually exclusive. To do this, we highlight a behavioral pattern observed in Experiments 1 and 2 that is consistent with the encoding accuracy account and then perform

Figure 3

Systematic VWM Errors Following "Similar" and "Dissimilar" Judgments



Note. (A) Signed response distributions for memory reports following "similar" and "dissimilar" judgments in the 45° probe condition. For illustration purposes, we plotted the proportion of responses for a given signed offset by calculating the mean response proportion across a 30° window centered at the offset. Positive offsets indicate memory errors toward the probe. The inset shows a close-up of the peak of each distribution. Shaded regions surrounding the distribution curve indicate within-subject standard errors of the mean (Cousineau, 2005). The vertical black and red (dark) dashed lines indicate the location of the target and the probe in the feature space, respectively, across trials. (B) Boxplots of the mean signed response error in each judgment condition with corresponding density distributions. Colored dots to the right of each boxplot indicate the mean error for a given participant. VWM = visual working memory; Exp. 1 = Experiment 1; Exp. 2 = Experiment 2. See the online article for the color version of this figure.
*** $p < .001$.

a proof of concept using simulations to show that this pattern is not inconsistent with memory replacement. Specifically, the encoding accuracy account predicts that trials where an identical probe (i.e., separated by 0° from the target) was correctly endorsed as the "same" should result in more precise memory reports, since perceived sameness should only be possible if participants encoded an accurate representation of the target. We show that memory reports were indeed more precise following correct "same" judgments than they were in the baseline condition, where poorly encoded target representations were not filtered in any way, and that these improvements in precision can occur even when the probe replaces the encoded target.

In the second prong of our analysis, we show that the encoding accuracy account cannot fully explain memory errors observed following inaccurate "same" judgments without the need for an additional mechanism. Using the same simulation procedure, we tested whether the distribution of memory errors observed following "same" judgments in the 15° probe condition could be recreated just by sampling memory reports from the baseline condition that were near the probe. In doing so, we found that filtering responses based on encoding accuracy was unable to recapitulate erroneous report patterns. Thus, in the following analyses, we show that encoding processes help determine the representational overlap that is perceived during perceptual comparisons, but these processes are not sufficient in explaining systematic memory errors that follow from comparisons.

Method

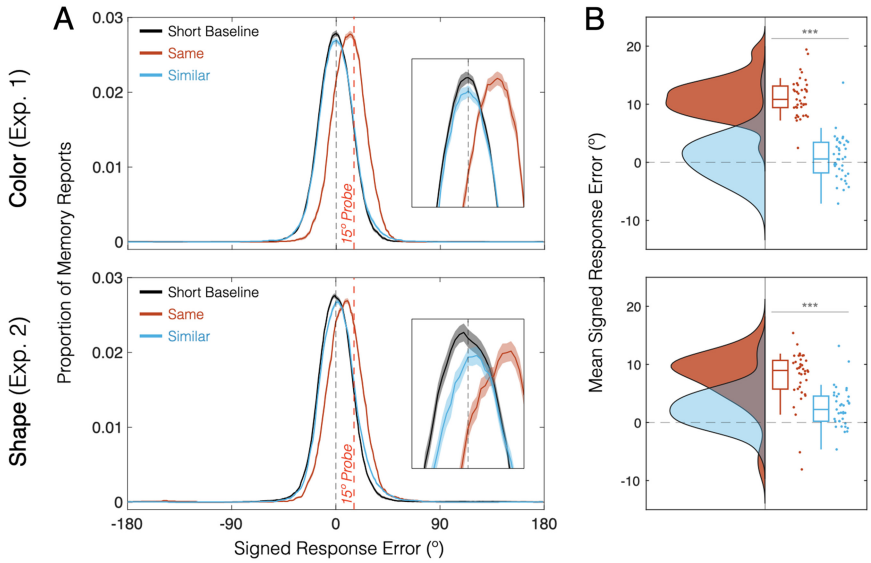
Simulation Analysis

In this approach, we implement the assumptions of the encoding accuracy account to simulate a collection of memory reports following "same" judgments that are based solely on noisy encoding processes. The crux of our logic is that memory errors determined at the time of encoding occur in all conditions, including the baseline conditions. Therefore, if the errors observed following perceived sameness solely reflect fluctuations in encoding accuracy, we should be able to re-produce the observed pattern of memory reports by sampling reports from the delay-matched short baseline condition that would have been near the location of the probe in the feature space (see Figure 5A and B for a visual illustration). Note that the delay between encoding and reporting in the short baseline condition was identical to the delay between encoding and comparisons in the experimental conditions, thereby allowing us to approximate the accuracy of the target at the time of the comparison using observed data.

To simulate a collection of memory reports that follow the logic of the encoding accuracy account, we sampled memory reports from the short baseline condition that would have been immediately surrounding the location of the hypothetical probe in the feature space in the experimental condition. The number of samples that were

Figure 4

Systematic VWM Errors Following "Same" and "Similar" Judgments



Note. (A) Signed response distributions for memory reports following "same" and "similar" judgments in the 15° probe condition. For illustration purposes, we plotted the proportion of responses for a given signed offset by calculating the mean response proportion across a 30° window centered at the offset. Positive offsets indicate memory errors toward the probe. The inset shows a close-up of the peak of each distribution. Shaded regions surrounding the distribution curve indicate within-subject standard errors of the mean (Cousineau, 2005). The vertical black and red (dark) dashed lines indicate the location of the target and the probe in the feature space, respectively, across trials. (B) Boxplots of the mean signed response error in each judgment condition with corresponding density distributions. Colored dots to the right of each boxplot indicate the mean error for a given participant. VWM = visual working memory; Exp. 1 = Experiment 1; Exp. 2 = Experiment 2. See the online article for the color version of this figure. *** $p < .001$.

drawn from the baseline condition was based on the number of "same" responses that were made by the participant.

For example, imagine a participant that endorsed 15°-offset probes to be identical ("same") on 30% of trials in the 15° probe condition. According to the encoding accuracy account, this would indicate that the participant encoded target θ as θ + 15° prior to the perceptual comparison on 30% of trials. To simulate this, we determined the range of response offsets surrounding 15° in the baseline condition that made up 30% of trials. We then multiplied the proportion of responses at each response offset within this range by the number of "same" responses that were made in the experimental condition. Participants that reported less than 10 "same" responses in the 15° probe condition were excluded from the respective simulation (color: $n = 8$ exclusions; shape: $n = 4$ exclusions). All participants reported more than 10 "same" responses in the 0° probe condition and were included in the respective simulation.

Results

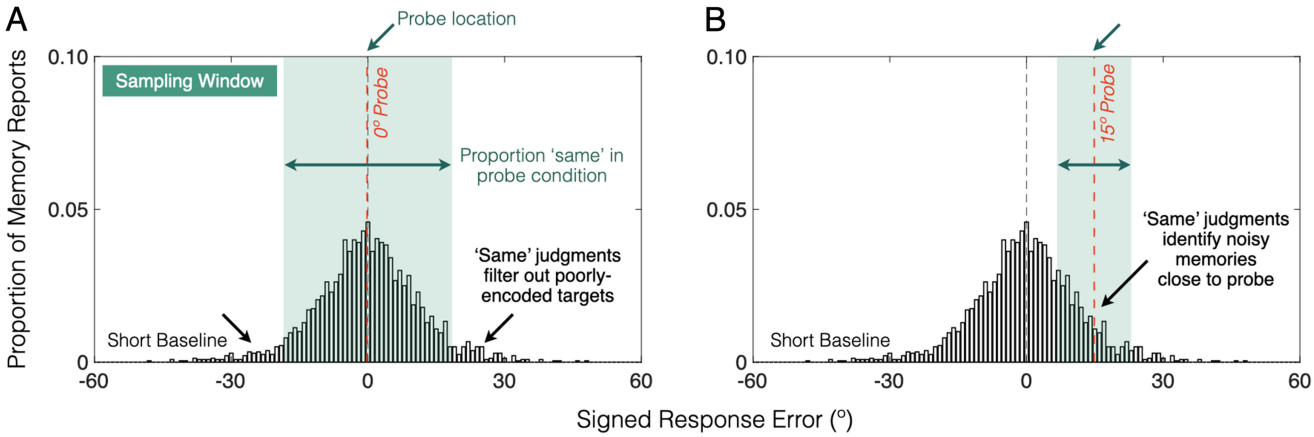
Improved Report Precision Is Compatible With Memory Replacement

Figure 6A shows the distribution of response errors that were observed following "same" judgments made to an identical probe stimulus. As can be seen, memory reports following these accurate "same" judgments were more tightly clustered around the zero-centered target than those made in the delay-matched short baseline condition. When we statistically compared the precision of the memory reports observed following accurate "same" judgments to those in the short baseline condition, we found evidence for higher precision in the prior (Figure 6B; color, $M = 0.019$, 95% CI [0.010, 0.027], $t(39) = 4.45$, $p < .001$, Cohen's $d = 0.70$, $BF_{10} = 3.40 \times 10^2$; shape, $M = 0.013$, [0.002, 0.025], $t(36) = 2.43$, $p = .020$, Cohen's $d = 0.40$, $BF_{10} = 2.32$). This pattern is consistent with a prediction made by the encoding accuracy account which asserts that perceived sameness depends on some minimum amount of representational overlap between the representations that effectively filtered out poorly encoded targets (Figure 5A).

To demonstrate that memory replacement is compatible with this finding, we conducted a simulation analysis to show that memory precision is higher in the identical probe condition even when the probe replaces the target. We simulated target representations for every trial by sampling responses from the short baseline condition that were near 0° (Figure 5A; see the Method section). Because the probe stimulus was physically identical to the target stimulus in these trials, we simulated independent encoding of the probe by duplicating the simulated targets and shuffling their order randomly. Finally, we simulated memory replacement using a parameter that determined how often a given memory report in the simulation was based on the probe rather than the target. Note that, in principle, the exact frequency

Figure 5

Simulating Memory Reports Based on Encoding Accuracy



Note. An encoding accuracy account of memory reports following perceived sameness may be simulated by sampling from a range of response errors surrounding the hypothetical probe in the short baseline condition where the sampled response is assumed to be close enough to the hypothetical probe that it is endorsed as being the "same." (A) When the probe is identical to the target, this process filters out poorly encoded target representations that are unlikely to be perceived as the "same." As a result, the number of responses along the tails of the response distribution is reduced, resulting in a greater proportion of responses clustered around the zero-centered target. (B) When the probe is different from the target, "same" judgments identify poorly encoded target representations that were near the probe prior to probe onset. As a result, response errors following "same" judgments are clustered around the probe, mimicking memory distortion. See the online article for the color version of this figure.

of memory replacement should not meaningfully influence the likelihood of increased precision except by chance alone. Figure 7 shows the results of the simulation when replacement was set to occur randomly in 50% of trials. Unsurprisingly, we found a clear increase in precision for the simulated responses relative to the observed responses in the short baseline condition. Additional iterations of this procedure confirmed that precision is improved regardless of how often replacement is set to occur in the simulation (see the online supplemental materials). Based on this proof of concept, we concluded that memory replacement is not inherently in conflict with the encoding accuracy account.

Errors Following Inaccurate "Same" Judgments Are Not Explained by Encoding Accuracy

Given that the encoding accuracy account was capable of explaining improvements in memory precision following identical probes without the need for memory updating, does this mean that memory errors following inaccurate "same" judgments can also be explained without memory updating? If so, we should be able to reconstruct the pattern of memory errors observed following "same" judgments in the 15° probe condition using the same simulation approach as before (Figure 5B; see the Method section).

Figure 8 shows the responses generated by the simulation plotted against the responses made by participants in the 15° probe condition. Simulated responses were shifted in the direction of the 15° probe, mimicking the shift that was present in participants' responses. However, because the simulated responses were necessarily sampled along the positive-going tail of the short baseline condition, there was a positive skew in the simulated responses (color: 0.53; shape: 0.32) that stood in stark contrast to the negative skew in the observed responses (color: −1.88; shape: −2.95). Visual inspection of the response distributions also revealed a clear difference in the precision of the simulated responses compared to those made by human observers. When we conducted Kolmogorov–Smirnov (KS) testing to assess whether the simulated and observed distributions were drawn from the same underlying distribution, we found evidence against this conclusion (color, $d = 0.18$, $p < .001$; shape, $d = 0.21$, $p < .001$). We then downsampled both distributions from 360 bins (1°/bin) to 72 bins (5°/bin) to counteract potential oversensitivity in our initial KS test. However, when we reconducted the KS test again on the downsampled data set, the difference persisted (color, $d = 0.16$, $p < .001$; shape, $d = 0.19$, $p < .001$). This suggests that the large, systematic report errors made following inaccurate "same" judgments were not merely tracking systematic differences in encoding accuracy and that another mechanism is required to explain these response patterns.

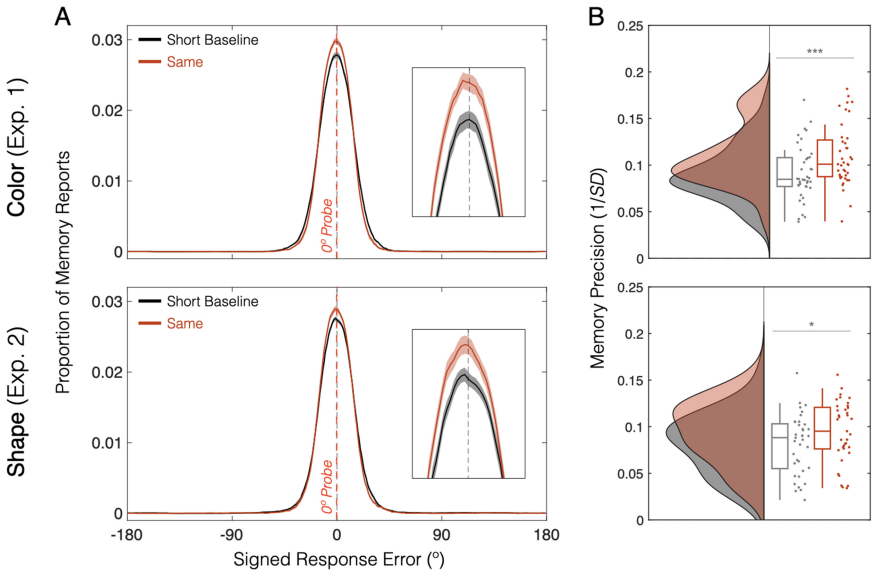
Discussion

We conducted simulations to address whether memory errors reported by participants were explained by the extent to which encoding accuracy differed systematically between subjective judgments. In doing so, we found that perceived sameness between a VWM and novel input requires some minimum amount of overlap between their representations. When the encoded target and novel input are physically identical, perceived sameness constrains the memory set to only include VWMs that were accurately encoded, resulting in precise reports of the target. Importantly, because the target and input were identical, reports are precise even if the input replaces the target in VWM. However, when the encoded target and novel input are different, constraining the memory set to only include targets that were noisily encoded to be like the input produces a pattern of memory reports that is meaningfully different from the one observed in behavior.

One may be tempted to explain the mismatch between the simulated and observed response distributions by highlighting other

Figure 6

Increased Report Precision Following Accurate "Same" Judgments



Note. (A) Signed response distributions for memory reports following "same" judgments in the identical probe condition and short baseline condition. For illustration purposes, we plotted the proportion of responses for a given signed offset by calculating the mean response proportion across a 30° window centered at the offset. Shaded regions surrounding the distribution curve indicate within-subject standard errors of the mean (Cousineau, 2005). The vertical black and red (dark) dashed lines indicated the location of the target and probe in the feature space, respectively, across trials. (B) Boxplots with corresponding density distributions depicting the mean response precision in the short baseline condition and following "same" judgments to an identical probe. Colored dots to the right of each boxplot indicate the mean precision for a given participant. Exp. 1 = Experiment 1; Exp. 2 = Experiment 2. See the online article for the color version of this figure.

* $p < .05$. *** $p < .001$.

potential sources of noise that were not accounted for in the simulation procedure. For example, memory reports in the delay-matched baseline condition that were used to simulate target representations may have been contaminated by motor and decision noise that would not have yet been present at the time of the comparison in the actual experiment. As a result, simulating responses from this baseline condition may not have fairly approximated the true accuracy of the target. While it is true that the current simulation procedure did not account for these sources of noise, it is highly unlikely that noise was responsible for the mismatch that was observed here. For one, both the observed and simulated response distributions each contain some amount of motor and decision noise since both are based on actual responses made by participants using an identical report procedure. Second, some results of the simulation simply cannot be explained by unsystematic sources of noise that increased the dispersion of responses in one condition more than the other. For example, opposite skewing directions in the observed and simulated response distributions can only be explained by a mechanism that is capable of altering the shape of the distribution. In the following section, we report the results of a computational modeling analysis where we demonstrate that the systematic variability in observers' memory reports following perceived similarity and perceived sameness are better approximated by two distinct memory updating mechanisms, namely representational integration (Fukuda et al., 2022; Saito et al., 2022) and representational replacement, respectively.

Modeling VWM Errors Following Perceptual Comparisons

To investigate the computational mechanisms responsible for VWM errors following perceived similarity and sameness, we identified two plausible models. First, we hypothesized that errors following "similar" judgments would be better fit by a joint density (JD) model which assumes that the target and probe representations are integrated to form a joint representation that is biased toward the probe (Bae et al., 2015; Fukuda et al., 2022; Saito et al., 2022). Second, we hypothesized that errors following "same" judgments would be better fit by a mixture density (MD) model which assumes that the probe representations probabilistically replace the target representations, leading to responses that appear biased but are underpinned by intact representations (Fukuda et al., 2022; Saito et al., 2022; see Bays et al., 2009 for a similar conceptualization). Together, we sought to illustrate that differences in the magnitude of memory errors between "similar" and "same" judgments reflect the consequences of qualitatively-distinct memory-updating mechanisms.

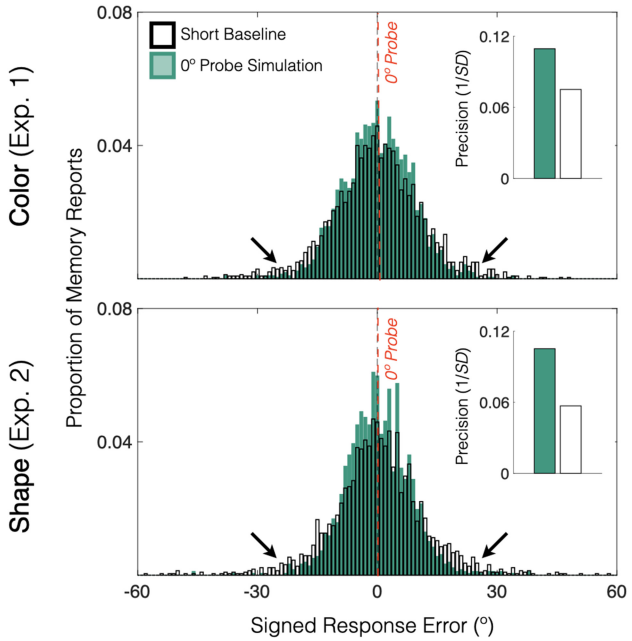
Method

JD Model

To account for systematic errors in VWM reports, the JD model assumes that participants' VWM representation of the target is

Figure 7

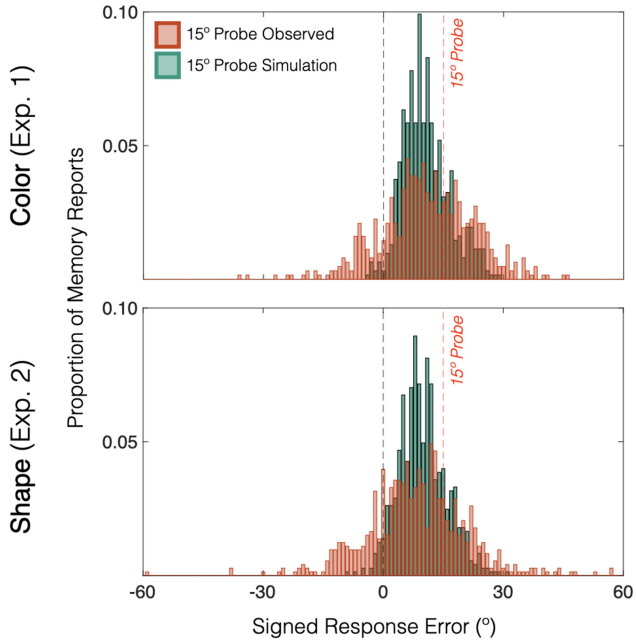
Improved Report Precision as a Function of Fluctuations in Encoding Accuracy



Note. Signed response distributions of simulated responses based on encoding accuracy and observed responses in the short baseline condition. In the simulated response distributions, responses along the tails are reduced compared to the baseline, resulting in a greater proportion of responses surrounding the target and identical probe. The inset bar chart shows the precision of the simulated responses compared to the observed baseline responses. In both color and shape, response precision was higher in the simulated response distribution despite the inclusion of probabilistic memory replacement, which was set to 50% frequency in the present figure. Exp. 1 = Experiment 1; Exp. 2 = Experiment 2. See the online article for the color version of this figure.

Figure 8

Systematic VWM Errors as a Function of Fluctuations in Encoding Accuracy



Note. Signed response distributions of simulated response errors and observed response errors following "same" judgments. VWM = visual working memory; Exp. 1 = Experiment 1; Exp. 2 = Experiment 2. See the online article for the color version of this figure.

integrated with the probe representation during perceptual comparisons. This integration process forms a JD distribution of the two representations that the participant randomly samples from at the time of the memory report. This model conceptualization is adopted from previous studies that demonstrated representational integration between visual representations (Bae et al., 2015; Fukuda et al., 2022). We construct and fit the JD model in three steps:

First, we construct the target representation by assuming a noisy representation (X_M) of the original target stimulus (S_M) that follows a von Mises distribution (φ) centered at the location of the target stimulus in the feature space with a given precision (κ_M).

$$p(X_M|S_M) = \phi(X_M|S_M + \mu, \kappa_M). \quad (1)$$

In the formula, the von Mises distribution contains parameters μ and κ_M. μ indicates the center of the distribution and is implemented to allow for shifting of the distribution relative to the actual stimulus. Here, μ is set to zero because we do not assume any systematic shift in the target representation that is initially encoded. κ_M indicates the concentration of the von Mises distribution, which corresponds to the precision of the target representation. κ_M was obtained by fitting a standard mixture model (Zhang & Luck, 2008) to the memory reports in the delay-matched short baseline condition where no probe was presented. This allowed us to base our precision estimate on observable behavior and increase the efficiency of our parameter search process.

Second, we construct the probe representation by assuming a noisy representation (X_P) of the probe stimulus (S_P) that again follows a von Mises distribution (φ) centered at the location of the probe stimulus in the feature space with a given precision (κ_P).

$$p(X_P|S_P) = \phi(X_P|S_P + \mu, \kappa_P). \quad (2)$$

μ again indicates the center of the von Mises distribution and is set to zero since we do not assume any systematic shift in the probe representation that is initially encoded. κ_P indicates the precision of the probe representation and was fit within the model as a free parameter. κ_P was set to vary in the parameter search process from 0.1 to 2 times κ_M in increments of 0.1 and from 2 to 10 times κ_M in increments of 0.5, yielding 36 possible precision estimates. That is, the precision of the probe was allowed to vary from 10% to 1,000% of the memory item. Separate analyses confirmed that varying the search range in our fitting procedure does not change the results of the model comparisons.

Third, we construct an integrated representation of the memory and probe items that is assumed to follow a joint density of the memory and probe distributions (X_{JD}).

$$p(X_{JD}|S_M, S_P) = \frac{p(X_M|S_M)p(X_P|S_P)}{\sum^{p}(X_M|S_M)p(X_P|S_P)}. \quad (3)$$

This JD distribution is constructed by a straightforward multiplication of the target (1) and probe (2) density functions. The four

parameters in the joint density function are identical to those in (1) and (2). Thus, we estimated one free parameter in the JD model (i.e., κ_P).

MD Model

To account for systematic errors in the VWM reports, the MD model assumes that participants' VWM representation of the target is sometimes replaced by the probe representation, such that individuals rely on the probe representation during the memory report. Thus, the model assumes that perceptual comparisons do not change the underlying memory representations, but instead change the likelihood that the probe representation will be used to represent the original target stimulus.

We construct and fit the MD model in three steps. Steps 1–2 are identical to the JD model. The third step is as follows:

$$p(X_{\text{Mix}}|S_M, S_P) = \alpha p(X_M|S_M) + (1 - \alpha)p(X_P|S_P) \quad (4)$$

α is used as a mixture parameter that estimates the proportion of trials where the memory report was based on the target representation (1). The remaining trials are assumed to be based on the probe representation (2) (i.e., $1-\alpha$). α was allowed to vary between 0 and 1 in increments of 0.05, yielding 21 possible mixture parameters. That is, the percentage of memory-based reports was allowed to vary from 0% (100% probe-based) to 100% (0% probe-based) in 5% increments. All other aspects of the MD model are identical to the JD model. Thus, we estimated two free parameters in the MD model (i.e., α and κ_P).

Procedure

We focused our model fitting for "same" responses on the 15° probe condition since there were too infrequent "same" responses in the 45° probe condition (see Figure 2). Likewise, we focused our model fitting for "similar" responses on the 45° probe condition where we found a reliable shift in the response distribution that was less apparent in the 15° probe condition (see Figure 3).

To limit contamination by trials where participants relied on guessing and other nonmnemonic response strategies (e.g., Pratte, 2019), we focused our model fitting on trials where participants reported being highly confident in their memory report. Given the limited number of confident "same" and "similar" judgments per subject ("same" color: $M = 13.80$ trials; "same" shape: $M = 16.65$ trials; "similar" color: $M = 14.35$ trials; "similar" shape: $M = 15.38$ trials), we maximized power in both of our model fitting procedures by aggregating responses across subjects and fitting the models to the group data. We also removed outlier trials that contained memory errors that were >2.5 *SD* above or below the mean response error of the group ("same" color: eight [1.45%] trials; "same" shape: nine [1.46%] trials; "similar" color: 16 [2.79%] trials; "similar" shape: 12 [2.11%] trials) since these trials can disproportionately skew formal model fit statistics (e.g., log likelihood; Huber, 2004). Indeed, when we reconducted our modeling with these outlier trials included, all model fits were weakened (see the online supplemental materials).

To construct a density distribution of the observed data, we used participants' signed response errors (in radians) to compute kernel density estimates at 10,000 equally spaced points along the circular space (i.e., −pi to pi; bin size = 0.0006 radians). We computed the same number of kernel density estimates for the best-fitting model distribution by reconstructing the predicted distribution using the parameters identified in our fitting procedure.

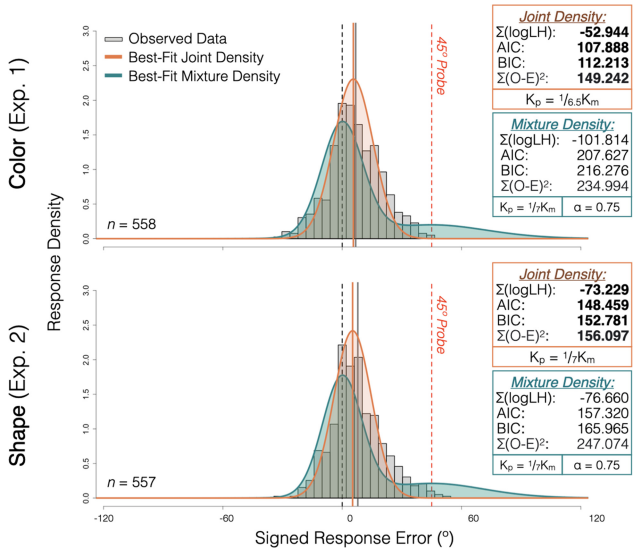
To find the best-fitting free parameters within each model, we calculated the sum of squared differences (i.e., Σ[Observed−Expected]^2 or sum of squared O-E) between the model density distribution that was constructed on each parameter search iteration and the observed density distribution and selected the parameter values that best minimized the difference between the distributions. We then compared the best-fitting JD and MD models using formal model fit statistics (i.e., sum of the log-likelihood or sum of the log LH, Akaike Information Criterion or AIC, Bayesian Information Criterion or BIC).

Results

First, we fit both models to errors observed following "similar" judgments in the 45° probe condition to test whether representational integration provided a better explanation for these errors than representational replacement. As can be seen in Figure 9, the bias observed across trials was accomplished by a positive shift in the central Gaussian toward the probe stimulus. Importantly, a positive shift in the central Gaussian can be produced by representational integration, since this process shifts the location of the target representation in the feature space, but cannot be produced by probabilistic replacement,

Figure 9

Computational Modeling of VWM Response Errors Following "Similar" Judgments



Note. Best-fitting JD and MD distributions for the color and shape experiments overlaid on the observed data. For visualization, the observed data in each experiment are plotted as a histogram with 90 bins (0.07 radians/bin). The *x*-axes are abbreviated to −120 to 120° surrounding the zero-centered target to emphasize the shape of the distributions. Vertical black and red (dark) dashed lines indicate the location of the target and probe stimuli in the feature space across trials, respectively. Vertical orange (dark) and gray solid lines indicate the mean response error in the best-fitting JD and observed data distributions, respectively. Formal model fit statistics and the sum of squared differences between the observed and best-fitting model data are reported with boldface values indicating the preferred model. Free parameters identified within the best-fitting models are reported below the fit statistics. VWM = visual working memory; Exp. 1 = Experiment 1; Exp. 2 = Experiment 2; JD = joint density; MD = mixture density. See the online article for the color version of this figure.

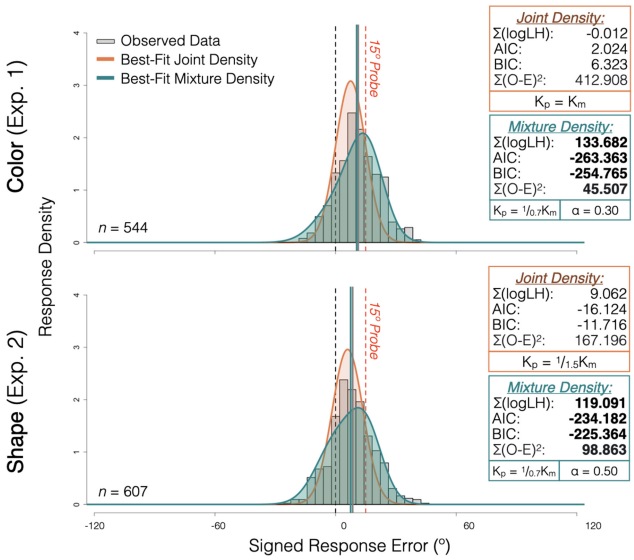
which creates a bimodal response pattern instead. There is no apparent evidence of bimodality in either of the observed response distributions. However, because the MD model assumes a bimodal response pattern, the best-fitting MD models were incapable of re-producing shifted central Gaussians and attempted to compensate for this by assuming that replacement occurred in only 25% of trials (i.e., alpha = .75) and that the responses based on the probe representation were widely distributed, producing long, positive-going tails in the distribution that was not present in the observed data. While the JD model did a superior job at capturing the positive shift in the central Gaussian, it did struggle to account for the skewing in the positive-going tail. This was likely due to fitting the model at the aggregate "super subject" level, which eliminated individual differences in VWM precision that may have influenced the magnitude of the bias. Note that previous studies that have fit the JD model at the individual subject level were able to reproduce this skewing in the probe-side tail (Fukuda et al., 2022). Finally, both the JD and MD models assumed an imprecise probe representation across trials, consistent with the central tenet of the encoding accuracy account which asserts that the subjective perception of overlap depends on some minimum amount of representational overlap. Because we fixed the precision of the target across trials in our model fitting by using the precision observed in the delay-matched baseline condition (see the Method section), encoding accuracy is accounted for in the probe representation instead.

The sum of squared differences measures in color and shape corroborated this qualitative assessment (color, JD = 149.242, MD = 234.994; shape, JD = 156.097, MD = 247.074), confirming that the shape of the best-fitting JD model more closely resembled the shape of the observed data than the best-fitting MD model. The sum of log-likelihood (color, JD = −52.944, MD = −101.814; shape, JD = −73.229, MD = −76.660), AIC (color, JD = 107.888, MD = 207.627; shape, JD = 148.459, MD = 157.320), and BIC (color, JD = 112.213, MD = 216.276; shape, JD = 152.781, MD = 165.965) measurements all unanimously preferred the JD model over the MD model. Thus, consistent with prior studies (Fukuda et al., 2022; Saito et al., 2022), we find that response errors following "similar" judgments were better explained by representational integration than probabilistic replacement.

We then moved to complement these initial findings by providing divergent evidence favoring the MD model in memory reports following "same" judgments in the 15° probe condition. Figure 10 shows the best-fitting JD and MD distributions along with the distribution of observed response errors following "same" judgments in the 15° probe condition. Unlike the 45° probe condition, the target and probe representations in the 15° probe condition were highly overlapping, resulting in a predicted MD distribution that was negatively skewed toward the probe rather than comprised of distinct bimodal peaks. This negative skewing in the MD distributions is consistent with the negative skewing present in the observed distributions (observed color: −1.90; observed shape: −2.93) and was critical in allowing the MD model to account for response errors that fell beyond the probe (i.e., errors > 15°). In contrast, the JD model struggled to account for response errors beyond 15° since integration necessarily produces a representation whose center falls between the target and the probe (i.e., 0° < center < 15°). We also noted that the JD and MD models now assumed probe precision to be roughly equal to or higher than target precision. Given the close physical proximity between the target and probe, it is reasonable that both representations could be encoded precisely, yet still considerably overlapping.

Figure 10

Computational Modeling of VWM Response Errors Following "Same" Judgments



Note. Best-fitting JD and MD distributions for the color and shape experiments overlaid on the observed data. For visualization, the observed data in each experiment are plotted as a histogram with 90 bins (0.07 radians/bin). The *x*-axes are abbreviated to −120 to 120° surrounding the zero-centered target to emphasize the shape of the distributions. Vertical black and red (dark) dashed lines indicate the location of the target and probe stimuli in the feature space across trials, respectively. Vertical teal and gray solid lines indicate the mean response error in the best-fitting MD and observed data distributions, respectively. Formal model fit statistics and the sum of squared differences between the observed and best-fitting model data are reported with boldface values indicating the preferred model. Free parameters identified within the best-fitting models are reported below the fit statistics. VWM = visual working memory; JD = joint density; MD = mixture density; Exp. 1 = Experiment 1; Exp. 2 = Experiment 2; AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion; LH = Likelihood; O-E = Observed minus Expected. See the online article for the color version of this figure.

Consistent with our qualitative assessment of the distributions, the sum of squared differences measures was lower in the MD model than in the JD model (color, JD = 412.908, MD = 45.507; shape, JD = 167.196, MD = 98.863). Formal model comparisons using the sum of log-likelihood (color, JD = −0.012, MD = 133.682; shape, JD = 9.062, MD = 119.091), AIC (color, JD = 2.024; MD = −263.363; shape, JD = −16.124, MD = −234.182), and BIC (color, JD = 6.323; MD = −254.765; shape, JD = −11.716, MD = −225.364) measurements all unanimously preferred the MD model over the JD model. Therefore, in conjunction with the modeling results for "similar" response errors, we find clear computational evidence that differences in the magnitude of response errors following "same" and "similar" judgments reflected the recruitment of qualitatively distinct memory updating mechanisms.

General Discussion

In the present experiments, we tested the prediction that using memories in perceptual comparisons can trigger distinct forms of

memory updating. Specifically, we predicted that perceived similarity between mnemonic and perceptual representations would result in representational integration, as demonstrated previously (Fukuda et al., 2022; Saito et al., 2022, 2023), and that perceived sameness would result in probabilistic replacement by the perceptual representation. We based this prediction on the fact that perceived sameness between sequential stimuli indicates a sense of shared identity that is fundamentally untrue for perceived similarity and allows individuals to rely on the more recent stimulus to facilitate their behavioral goals (e.g., Schacter et al., 2011). This framework is also consistent with longstanding literature showing that everyday memory errors can occur as a result of misattributing new perceptual details to a prior experience, especially when individuals fail to detect differences between their memory and these novel details (Butler & Loftus, 2018; Greene et al., 1982; Loftus, 1992; Thomas et al., 2010; Tousignant et al., 1986; Zaragoza & Lane, 1994). In these cases, novel details do not bias existing memories, but replace them.

In accordance with our predictions, we found that memory reports were shifted towards the probe following "same" and "similar" judgments and that this shift was larger when the memory and probe were perceived to be the same. We conducted simulations to test whether these sizable errors following "same" judgments reflected systematic differences in the quality of initial memory encoding rather than updating driven by the perceptual comparison (see Fukuda et al., 2022 for a direct investigation of this issue in "similar" judgments). These simulations revealed that perceiving sameness depends on some minimum amount of representational overlap between the memory and probe that can explain higher report precision following accurate "same" judgments but cannot fully capture errors that occur following inaccurate "same" judgments. We then used computational modeling to try and map memory errors following "same" and "similar" judgments to two qualitatively distinct mechanisms that have been described in prior studies (Bae et al., 2015; Bays et al., 2009; Fukuda et al., 2022; Saito et al., 2022). In doing so, we showed that systematic response errors following "same" judgments were better captured by representational replacement than representational integration, while the vice versa was true for errors following "similar" judgments. Together, these findings help to advance the burgeoning perspective that perceptual comparisons underlie memory distortions that are observed in everyday life, including those that are not amenable to a biased account.

Upon initial consideration, it may appear surprising that changing the judgment during a perceptual comparison changes the form of memory updating that transpires. From a signal detection perspective, perceived sameness and similarity are not fixed categorical states along the continuum of representational overlap. Rather, the shift from perceived similarity to sameness is assumed to occur when the amount of representational overlap between the memory and percept exceeds a decision criterion that is set by the observer to fulfill the demands of the comparison (Morrell et al., 2002; Wixted, 2007). Decision criteria are known to be sensitive to changes in task context, such that observers can become more liberal or conservative in what they endorse as being the same. In principle, this could mean that the decision (i.e., judgment) made during a perceptual comparison could change independently of the psychological experience that drives memory updating (i.e., perceived overlap). In the present study, because we manipulated the likelihood of perceived sameness and similarity by changing the physical similarity between the target and the probe on each trial, changes in the judgment likely coincided closely with changes in the perceived overlap. However, other studies have shown that the mnemonic consequences of perceived sameness are observed even in cases when only decision criteria are manipulated. For example, studies of eyewitness testimony have shown that observers can be made more likely to falsely identify an innocent suspect within a lineup just by manipulating how well the other "fillers" in the lineup resemble the perpetrator (Colloff et al., 2016; Fitzgerald et al., 2013; Wixted & Mickes, 2014). In these circumstances, even though the likelihood of perceiving sameness is inflated by a decisional bias and not by the amount of representational overlap, eyewitnesses tend to confidently sustain these false identifications across time, suggesting that observers still experienced perceived sameness, which resulted in memory replacement (Roediger & DeSoto, 2015; Roediger et al., 2012; Wixted et al., 2015). Thus, while judgments made during perceptual comparisons can change independently of representational overlap, the correspondence between the judgment and the underlying psychological experience may be preserved.

However, the case for perceptual comparisons as a cognitive mechanism requires more investigation. For example, it is unknown whether judgments made during perceptual comparisons must be explicit for distinct memory updating to emerge. In previous studies of comparison-induced distortion, participants were always asked to endorse whether the percept was similar or dissimilar to the target memory (Fukuda et al., 2022; Saito et al., 2022, 2023). In those investigations, systematic errors observed following comparisons were always explained by a modulation in naturally occurring memory biases following perceived similarity. Is it possible that omitting the option to endorse the probe as the "same" in those studies reduced or eliminated the likelihood of perceived sameness and, as a result, memory replacement? The answer to this question could depend on whether systematic report errors following "same" judgments reflect bona fide changes in the memory representation at the time of the comparison or changes in decisional processes at the time of the memory report. Previous work has shown that report biases following "similar" judgments cannot be explained by trivial report strategies in which observers intentionally fine-tune their responses toward the probe (Saito et al., 2023; see also Chunharas et al., 2022). Nonetheless, it could be the case that explicit judgments influence observers' weighting of the probe stimulus as a decisional prior when memory representations are read out as behavioral reports (see, e.g., Brady et al., 2018; Hemmer & Steyvers, 2009; Honig et al., 2020; Huttenlocher et al., 2000). The most direct approach for addressing each of these questions will be to incorporate a neuroimaging method that would allow researchers to measure perceived overlap implicitly while tracking the contents of VWM before, during, and after probe perception (e.g., Bae, 2021; Harrison & Tong, 2009; Rademaker et al., 2019; Serences et al., 2009).

A comprehensive framework for predicting perceived overlap during perceptual comparisons also remains wanting. Across multiple experiments, we find that even when the physical distance between the target and probe is held constant, participants' judgments during perceptual comparisons vary from trial to trial (Fukuda et al., 2022; Saito et al., 2022). It is enticing to assume that this variability is determined exclusively by the precision of the memory representation at the time of the comparison. For example, in a recent study, researchers were able to predict the occurrence of attraction and repulsion biases between VWM representations based on trial-wise manipulations of memory precision and participants' subjective confidence (Lively et al., 2021). However, we have shown that memory errors following perceptual comparisons are

reliably larger than those following perceptual interference and simultaneous maintenance, even when baseline precision is matched between tasks and participants are highly confident in their memory reports (Saito et al., 2022). Moreover, we found in the present experiments that memory errors following "same" judgments could not be readily explained by imprecise target memories that were encoded to be like the probe prior to comparisons. This is not to say that memory fidelity does not influence the perceived overlap between representations, but that variability in memory fidelity alone is not sufficient in accounting for the memory errors that are associated with a given subjective judgment.

Then what other processes can plausibly contribute to the observed variability in perceptual comparisons? One possibility is that judgments during perceptual comparisons are influenced by general cognitive states that carry over from prior behavior. Multiple studies have demonstrated that performing recognition judgments can bias the memory system in or out of a discriminatory state and that this change in state can influence subsequent recognition judgments (e.g., Duncan et al., 2012; Patil & Duncan, 2018). Interestingly, these processing biases are shown to be especially influential when memories and perceptual inputs are similar, but not identical. A change in one's cognitive state may therefore act as an additional impetus that can lead to different subjective judgments during perceptual comparisons, even when memory precision and stimulus distance are comparable. Future work should investigate the serial nature of perceptual comparisons by carefully manipulating the overlap between visual representations across trials in an attempt to exploit a possible carryover effect.

In addition to providing a better predictive framework of perceived overlap, it will be necessary to formalize the observed memory updating processes in a more precise manner. In the present study, JD and MD models were constructed to compare the relative fit between two dissociable memory updating mechanisms to the systematic errors observed in participants' memory reports. We leveraged the same model formalizations that have been used previously to elucidate similarity-induced memory biases (Fukuda et al., 2022; Saito et al., 2022). However, some findings in the present study suggest that there may have been additional interactions between the target and probe that were not accounted for in these parsimonious versions of the models. For example, the distribution of memory reports observed following "same" judgments in the 15° probe condition was not centered at 15° in the feature space, despite the MD model's assumption that perceived sameness triggers replacement by the probe. This may be because the current implementation of the MD model does not account for biases in the perception of the probe that may have been induced by the target before memory replacement occurred. Previous studies have shown that incoming visual percepts can be proactively biased by previous percepts that are maintained in VWM (Kang et al., 2011; Olkkonen & Allred, 2014; Scocchia et al., 2013) and those that have already been discarded (Bae & Luck, 2019, 2020; Fischer & Whitney, 2014; Fritsche et al., 2017). In the present experiments, actively maintaining the target in VWM may have resulted in a proactive bias in participants' perception of the probe that caused it to appear more similar to the target than it actually was. As a result, when the probe replaced the target in VWM, the resulting memory report was always slightly offset from the true location of the probe in the direction of the target. This proactive influence on the probe representation may have even caused the memory representation to replace the probe representation in some trials. Future studies should seek to address the role of serial dependence in the present paradigm by measuring proactive biases and then using those empirical estimates to fit the models accordingly. In doing so, it may be revealed that the frequency of memory replacement was underestimated in the present study by fitting the models with the physical location of the probe rather than its perceived location.

Some may also wonder whether the outcomes of the model comparisons reported here depend on the assumed psychophysical properties of VWM. Here, JD and MD models were constructed following the assumptions of conventional models of VWM which state that the psychophysical similarity between stimuli in a given feature space is linearly related to their physical similarity (e.g., Zhang & Luck, 2008). However, recent research has challenged this conventional modeling assumption by demonstrating that psychophysical similarity is actually nonlinearly related to physical similarity (Schurgin et al., 2020). While nonlinearity in psychophysical scaling does not necessarily invalidate the model comparisons reported here, it is important to consider what role, if any, psychophysical scaling plays in representational integration and replacement. In the online supplemental materials, we report a replication of our model comparisons for color stimuli in Experiment 1 using the same psychophysical scaling properties and assumptions made by the target confusability competition (TCC) model that was proposed by Schurgin et al. (2020) to account for nonlinearity in the global similarity structure (but see also, Oberauer, 2023; Tomić & Bays, 2022). The results of the model comparison were not only replicated, but the TCC formulation of the JD and MD models produced best-fitting distributions that were very similar to those produced here. This suggests that, despite meaningful theoretical differences between the TCC model and conventional models of VWM, both frameworks provide a common conclusion regarding the memory updating processes tied to perceived sameness and similarity. Future work should seek to further extend these findings by replicating the model comparisons reported in Experiment 2 after precisely mapping the psychophysical scaling within the shape space.

Lastly, it will be useful for future work to build on these findings by testing the representational boundaries that perceptual comparisons operate within. While the distorting effect of perceptual comparisons has been illustrated in simple visual features (Fukuda et al., 2022; Saito et al., 2022), real-world objects (Saito et al., 2023), and human faces (Teoh et al., 2021; see also Plummer et al., 2021), no study has tested the effect of performing perceptual comparisons between event representations that are perceived in different sensory modalities. An overwhelming majority of studies on the misinformation effect have implemented paradigms in which false details about a prior *visual* experience are provided *verbally* to participants (see Loftus, 2005 for review). In the seminal experiment conducted by Loftus et al. (1978), postevent misinformation that was read by participants led to reliable misremembering of a street sign that they saw during a slide show depiction of an automobile accident. While studies like Loftus et al. (1978) confirm that memory misattribution can occur for individual objects embedded within events, it remains unclear how perceptual comparisons are executed when stimuli are drawn from different modalities. One possibility is that individuals form a parallel visual representation of verbal postevent inputs in order to facilitate more direct comparisons with their visual memory of the event. Researchers have shown that engaging in the visual imagery of a verbal stimulus can lead to more false memories than

when imagery is not used (e.g., Gonsalves & Paller, 2000; Gonsalves et al., 2004). Future work should examine these cross-modal effects more directly by incorporating neuroimaging techniques that can characterize the representational state of memories and perceptual inputs.

Constraints on Generality

In addition to the limitations described above, it is critical that future work clarify the generality of these findings to other human populations and other types of stimuli. In the present study, we show that young adults' working memory for a visual stimulus undergo distinct types of updating as a consequence of performing perceptual comparisons with new visual inputs. Given well-established age-related declines in WM performance (e.g., Peich et al., 2013; Pertzov et al., 2015), future studies should test whether our results extend to older adult populations as well. Such investigations should help to further elucidate the link between WM performance and the two types of memory updating described here. Additionally, future studies should examine how perceptual comparisons influence other types of memory representations beyond visual working memories, such as episodic memories and those from other sensory modalities (e.g., auditory memories). As mentioned above, such studies will provide critical observations to evaluate the role of perceptual comparisons as a domain-general mechanism of memory updating.

Conclusion

We used a delayed-estimation paradigm to show that comparing visual memories with new visual inputs can result in different types of memory updating. Across behavior, simulations, and computational modeling, we show that these distinct patterns of memory updating depend on the distinct judgments that individuals make during comparisons. These findings corroborate and extend mounting evidence that the use of a VWM in task-relevant behavior holds significant implications for the accuracy of the memory thereafter.

References

Aizpurua, A., Garcia-Bajos, E., & Migueles, M. (2009). Memory for actions of an event: Older and younger adults compared. *The Journal of General Psychology: Experimental, Psychological, and Comparative Psychology*, *136*(4), 428–441. https://doi.org/10.1080/00221300903269816

Bae, G. Y. (2021). Neural evidence for categorical biases in location and orientation representations in a working memory task. *NeuroImage*, *240*, 1–13. https://doi.org/10.1016/j.neuroimage.2021.118366

Bae, G. Y., & Luck, S. J. (2019). Reactivation of previous experiences in a working memory task. *Psychological Science*, *30*(4), 587–595. https://doi.org/10.1177/0956797619830398

Bae, G. Y., & Luck, S. J. (2020). Serial dependence in vision: Merely encoding the previous-trial target is not enough. *Psychonomic Bulletin & Review*, *27*(2), 293–300. https://doi.org/10.3758/s13423-019-01678-7

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

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

Brady, T. F., Schacter, D. L., & 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

Brainerd, C. J., & Reyna, V. F. (2005). *The science of false memory*. Oxford University Press.

Braun, K. A., & Loftus, E. F. (1998). Advertising's misinformation effect. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *12*(6), 569–591. https://doi.org/10.1002/(SICI)1099-0720(1998120)12:6<569::AID-ACP539>3.0.CO;2-E

Butler, B. J., & Loftus, E. F. (2018). Discrepancy detection in the retrieval-enhanced suggestibility paradigm. *Memory*, *26*(4), 483–492. https://doi.org/10.1080/09658211.2017.1371193

Chunharas, C., Rademaker, R. L., Brady, T. F., & Serences, J. T. (2022). An adaptive perspective on visual working memory distortions. *Journal of Experimental Psychology: General*, *151*(10), 2300–2323. https://doi.org/10.1037/xge0001191

Colloff, M. F., Wade, K. A., & Strange, D. (2016). Unfair lineups make witnesses more likely to confuse innocent and guilty suspects. *Psychological Science*, *27*(9), 1227–1239. https://doi.org/10.1177/0956797616655789

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology*, *1*(1), 42–45. https://doi.org/10.20982/tqmp.01.1.p042

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*(1), 87–114. https://doi.org/10.1017/S0140525X01003922

Duncan, K., Sadanand, A., & Davachi, L. (2012). Memory's penumbra: Episodic memory decisions induce lingering mnemonic biases. *Science*, *337*(6093), 485–487. https://doi.org/10.1126/science.1221936

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Fischer, J., & Whitney, D. (2014). Serial dependence in visual perception. *Nature Neuroscience*, *17*(5), 738–743. https://doi.org/10.1038/nn.3689

Fitzgerald, R. J., Price, H. L., Oriet, C., & Charman, S. D. (2013). The effect of suspect-filler similarity on eyewitness identification decisions: A meta-analysis. *Psychology, Public Policy, and Law*, *19*(2), 151–164. https://doi.org/10.1037/a0030618

Fritsche, M., Mostert, P., & de Lange, F. P. (2017). Opposite effects of recent history on perception and decision. *Current Biology*, *27*(4), 590–595. https://doi.org/10.1016/j.cub.2017.01.006

Fukuda, K., Pereira, A. E., Saito, J. M., Tang, T. Y., Tsubomi, H., & Bae, G. Y. (2022). Working memory content is distorted by its use in perceptual comparisons. *Psychological Science*, *33*(5), 816–829. https://doi.org/10.1177/09567976211055375

Fukuda, K., & Woodman, G. F. (2017). Visual working memory buffers information retrieved from visual long-term memory. *Proceedings of the National Academy of Sciences*, *114*(20), 5306–5311. https://doi.org/10.1073/pnas.1617874114

Gonsalves, B., & Paller, K. A. (2000). Neural events that underlie remembering something that never happened. *Nature Neuroscience*, *3*(12), 1316–1321. https://doi.org/10.1038/81851

Gonsalves, B., Reber, P. J., Gitelman, D. R., Parrish, T. B., Mesulam, M. M., & Paller, K. A. (2004). Neural evidence that vivid imagining can lead to false remembering. *Psychological Science*, *15*(10), 655–660. https://doi.org/10.1111/j.0956-7976.2004.00736.x

Greene, E., Flynn, M. S., & Loftus, E. F. (1982). Inducing resistance to misleading information. *Journal of Verbal Learning & Verbal Behavior*, *21*(2), 207–219. https://doi.org/10.1016/S0022-5371(82)90571-0

Harrison, S. A., & Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. *Nature*, *458*(7238), 632–635. https://doi.org/10.1038/nature07832

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

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

Huber, P. J. (2004). *Robust statistics* (Vol. 523). John Wiley & Sons.

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

Kang, M. S., Hong, S. W., Blake, R., & Woodman, G. F. (2011). Visual working memory contaminates perception. *Psychonomic Bulletin & Review*, *18*(5), 860–869. https://doi.org/10.3758/s13423-011-0126-5

Kleiner, M., Brainard, D. H., & Pelli, D. G. (2007). What's new in psychtoolbox-3? *Perception*, *36*(14), 1–16. ECVP Abstract Supplement. https://doi.org/10.1177/03010066070360S10

Li, A. Y., Liang, J. C., Lee, A. C., & Barense, M. D. (2020). The validated circular shape space: Quantifying the visual similarity of shape. *Journal of Experimental Psychology: General*, *149*(5), 949–966. https://doi.org/10.1037/xge0000693

Lively, Z., Robinson, M. M., & Benjamin, A. S. (2021). Memory fidelity reveals qualitative changes in interactions between items in visual working memory. *Psychological Science*, *32*(9), 1426–1441. https://doi.org/10.1177/0956797621997367

Loftus, E. F. (1992). When a lie becomes memory's truth: Memory distortion after exposure to misinformation. *Current Directions in Psychological Science*, *1*(4), 121–123. https://doi.org/10.1111/1467-8721.ep10769035

Loftus, E. F. (2005). Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learning & Memory*, *12*(4), 361–366. https://doi.org/10.1101/lm.94705

Loftus, E. F., Miller, D. G., & Burns, H. J. (1978). Semantic integration of verbal information into a visual memory. *Journal of Experimental Psychology: Human Learning and Memory*, *4*(1), 19–31. https://doi.org/10.1037/0278-7393.4.1.19

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

Mitchell, K. J., & Johnson, M. K. (2009). Source monitoring 15 years later: What have we learned from fMRI about the neural mechanisms of source memory? *Psychological Bulletin*, *135*(4), 638–677. https://doi.org/10.1037/a0015849

Morrell, H. E. R., Gaitan, S., & Wixted, J. T. (2002). On the nature of the decision axis in signal-detection-based models of recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(6), 1095–1110. https://doi.org/10.1037/0278-7393.28.6.1095

Oberauer, K. (2023). Measurement models for visual working memory—A factorial model comparison. *Psychological Review*, *130*(3), 841–852. https://doi.org/10.1037/rev0000328

Olkkonen, M., & Allred, S. R. (2014). Short-term memory affects color perception in context. *PLoS ONE*, *9*(1), Article e86488. https://doi.org/10.1371/journal.pone.0086488

Patil, A., & Duncan, K. (2018). Lingering cognitive states shape fundamental mnemonic abilities. *Psychological Science*, *29*(1), 45–55. https://doi.org/10.1177/0956797617728592

Peich, M.-C., Husain, M., & Bays, P. M. (2013). Age-related decline of precision and binding in visual working memory. *Psychology and Aging*, *28*(3), 729–743. https://doi.org/10.1037/a0033236

Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1–2), 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Pertzov, Y., Heider, M., Liang, Y., & Husain, M. (2015). Effects of healthy ageing on precision and binding of object location in visual short term memory. *Psychology and Aging*, *30*(1), 26–35. https://doi.org/10.1037/a0038396

Plummer, M., Hellerstedt, R., Gibson, S., Simons, J., & Bergstrom, Z. M. (2021). *Active Recognition Attempts Induce Updating of Face Memories*. *PsyArXiv*. https://doi.org/10.31234/osf.io/63qnj

Pratte, M. S. (2019). Swap errors in spatial working memory are guesses. *Psychonomic Bulletin & Review*, *26*(3), 958–966. https://doi.org/10.3758/s13423-018-1524-8

Qualtrics Inc. (2020). *Qualtrics: An online tool for creating and distributing surveys*. https://www.qualtrics.com/

Rademaker, R. L., Bloem, I. M., De Weerd, P., & Sack, A. T. (2015). The impact of interference on short-term memory for visual orientation. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(6), 1650–1665. https://doi.org/10.1037/xhp0000110

Rademaker, R. L., Chunharas, C., & Serences, J. T. (2019). Coexisting representations of sensory and mnemonic information in human visual cortex. *Nature Neuroscience*, *22*(8), 1336–1344. https://doi.org/10.1038/s41593-019-0428-x

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Roediger, H. L., III, & DeSoto, K. A. (2015). Understanding the relation between confidence and accuracy in reports from memory. In D. S. Lindsay (Ed.), C. M. Kelley (Trans.), A. P. Yonelinas, & H. L. Roediger, II (Eds.), *Remembering: Attributions, processes, and control in human memory: Essays in honor of Larry Jacoby* (pp. 347–367). Psychology Press.

Roediger, H. L., III, Wixted, J. H., & DeSoto, K. A. (2012). The curious complexity between confidence and accuracy in reports from memory. In L. Nadel & W. P. Sinnott-Armstrong (Eds.), *Memory and law* (pp. 84–118). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199920754.003.0004

Saito, J. M., Duncan, K., & Fukuda, K. (2023). Comparing visual memories to similar visual inputs risks lasting memory distortion. *Journal of Experimental Psychology: General*, *152*(8), 2318–2330. https://doi.org/10.1037/xge0001400

Saito, J. M., Kolisnyk, M., & Fukuda, K. (2022). Perceptual comparisons modulate memory biases induced by new visual inputs. *Psychonomic Bulletin & Review*, *30*(1), 291–302. https://doi.org/10.3758/s13423-022-02133-w

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

Schurgin, M. W., Wixted, J. T., & Brady, T. F. (2020). Psychophysical scaling reveals a unified theory of visual memory strength. *Nature Human Behaviour*, *4*(11), 1156–1172. https://doi.org/10.1038/s41562-020-00938-0

Scocchia, L., Valsecchi, M., Gegenfurtner, K. R., & Triesch, J. (2013). Visual working memory contents bias ambiguous structure from motion perception. *PLoS ONE*, *8*(3), Article e59217. https://doi.org/10.1371/journal.pone.0059217

Scotti, P. S., Hong, Y., Leber, A. B., & Golomb, J. D. (2021). Visual working memory items drift apart due to active, not passive, maintenance. *Journal of Experimental Psychology: General*, *150*(12), 2506–2524. https://doi.org/10.1037/xge0000890

Serences, J. T., Ester, E. F., Vogel, E. K., & Awh, E. (2009). Stimulus-specific delay activity in human primary visual cortex. *Psychological Science*, *20*(2), 207–214. https://doi.org/10.1111/j.1467-9280.2009.02276.x

Steblay, N. K., & Dysart, J. E. (2016). Repeated eyewitness identification procedures with the same suspect. *Journal of Applied Research in Memory and Cognition*, *5*(3), 284–289. https://doi.org/10.1016/j.jarmac.2016.06.010

Sun, S. Z., Fidalgo, C., Barense, M. D., Lee, A. C. H., Cant, J. S., & Ferber, S. (2017). Erasing and blurring memories: The differential impact of interference on separate aspects of forgetting. *Journal of Experimental Psychology: General*, *146*(11), 1606–1630. https://doi.org/10.1037/xge0000359

Sutterer, D. W., Foster, J. J., Serences, J. T., Vogel, E. K., & Awh, E. (2019). Alpha-band oscillations track the retrieval of precise spatial representations from long-term memory. *Journal of Neurophysiology*, *122*(2), 539–551. https://doi.org/10.1152/jn.00268.2019

Teng, C., & Kravitz, D. J. (2019). Visual working memory directly alters perception. *Nature Human Behaviour*, *3*(8), 827–836. https://doi.org/10.1038/s41562-019-0640-4

Teoh, Y. J., Khan, S., Yeo, Y., Saito, J. M., & Fukuda, K. (2021). *Comparisons with similar faces induce lasting distortions in face memories*. Poster presented at the Annual Meeting of the Vision Sciences Society, St. Petersburg, FL.

The Mathworks Inc. (2020). *MATLAB* (Version 9.8.0 (R2020a)). https://www.mathworks.com/

Thomas, A. K., Bulevich, J. B., & Chan, J. C. K. (2010). Testing promotes eyewitness accuracy with a warning: Implications for retrieval enhanced suggestibility. *Journal of Memory and Language*, *63*(2), 149–157. https://doi.org/10.1016/j.jml.2010.04.004

Tomić, I., & Bays, P. M. (2022). Perceptual similarity judgments do not predict the distribution of errors in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. Advance online publication. https://doi.org/10.1037/xlm0001172

Tousignant, J. P., Hall, D., & Loftus, E. F. (1986). Discrepancy detection and vulnerability to misleading postevent information. *Memory & Cognition*, *14*(4), 329–338. https://doi.org/10.3758/BF03202511

Vo, V. A., Sutterer, D. W., Foster, J. J., Sprague, T. C., Awh, E., & Serences, J. T. (2022). Shared representational formats for information maintained in working memory and information retrieved from long-term memory. *Cerebral Cortex*, *32*(5), 1077–1092. https://doi.org/10.1093/cercor/bhab267

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2001). Storage of features, conjunctions, and objects in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *27*(1), 92–114. https://doi.org/10.1037/0096-1523.27.1.92

Wixted, J. T. (2007). Dual-process theory and signal-detection theory of recognition memory. *Psychological Review*, *114*(1), 152–176. https://doi.org/10.1037/0033-295X.114.1.152

Wixted, J. T., & Mickes, L. (2014). A signal-detection-based diagnostic-feature-detection model of eyewitness identification. *Psychological Review*, *121*(2), 262–276. https://doi.org/10.1037/a0035940

Wixted, J. T., Mickes, L., Clark, S. E., Gronlund, S. D., & Roediger, H. L., III. (2015). Initial eyewitness confidence reliably predicts eyewitness identification accuracy. *American Psychologist*, *70*(6), 515–526. https://doi.org/10.1037/a0039510

Wixted, J. T., Mickes, L., Dunn, J. C., Clark, S. E., & Wells, W. (2016). Estimating the reliability of eyewitness identifications from police lineups. *Proceedings of the National Academy of Sciences*, *113*(2), 304–309. https://doi.org/10.1073/pnas.1516814112

Zaragoza, M. S., & Lane, S. M. (1994). Source misattributions and the suggestibility of eyewitness memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(4), 934–945. https://doi.org/10.1037/0278-7393.20.4.934

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

Zoom Video Communications Inc. (2020). *Zoom meetings* (Version 5.14.10) [Computer software]. https://www.zoom.us/

Received October 11, 2022
Revision received June 27, 2023
Accepted July 6, 2023 ■