Supplementary Material

for

Comparing Visual Memories to Similar Visual Input Risks Lasting Memory Distortion

Joseph M. Saito[1], Katherine Duncan[1], & Keisuke Fukuda[1,2]

[1]University of Toronto, [2]University of Toronto Mississauga

Contents:

1 Memory bias in Experiment 1 for all confidence levels

2 Experiment S1: Report biases following perceptual comparisons in a 2AFC paradigm

3 Memory bias in Experiment 2 for all confidence levels

1 Memory bias in Experiment 1 for all confidence levels

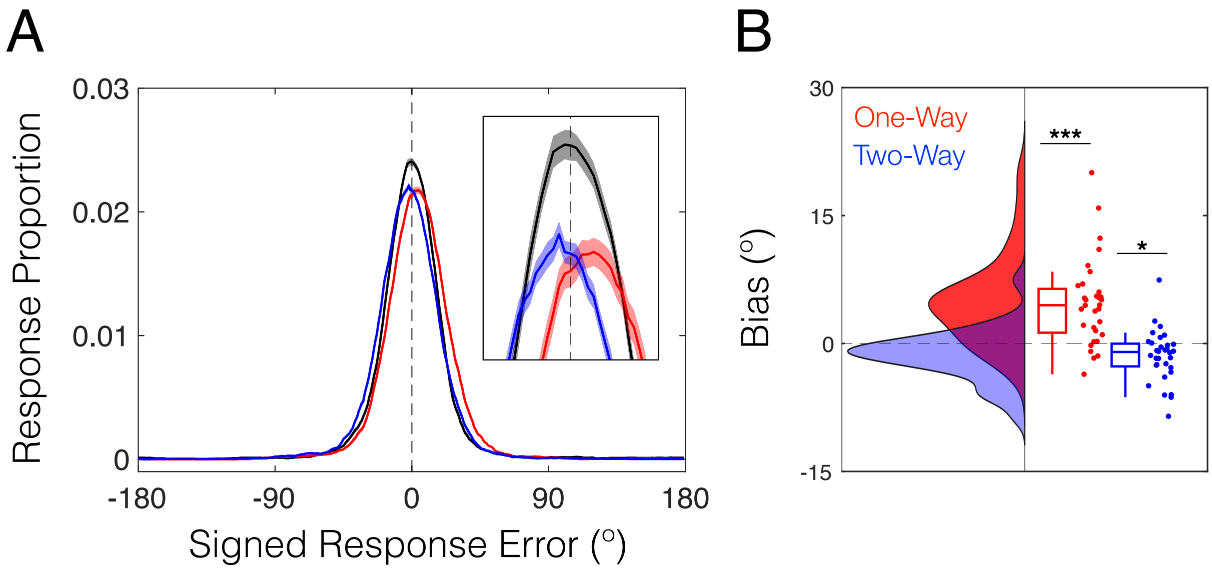

Figure S1. *Experiment 1 Results with All Trials Included.*

(A) Signed response distributions for the baseline and experimental conditions across all trials. For demonstration purposes, we plotted the proportion of responses for a given signed offset value by calculating the mean response proportion across a 30° window centered at the offset value. Positive offsets indicate memory bias towards the first similar 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 dashed line indicates the location of the target. (B) Boxplots of the mean signed response error (i.e., bias) in each experimental condition. Positive values indicate memory bias towards the first similar probe. The horizontal line inside each boxplot indicates the median bias across participants. Colored dots to the right of each boxplot indicate the bias for a given participant with corresponding density distributions shown on the left-hand panel of the figure. * $p < 0.05$, *** $p < 0.001$

Figure S1A shows the result of the main analyses when trials from all confidence levels were included. Here again, we observed a shift in the *one-way* response distribution that reflected reliable attraction towards the first similar probe (**Figure S1B**; $M = 4.72°$, 95% CI [2.90, 6.55°], $t(31) = 5.28$, $p < 0.001$, *Cohen's* $d = 0.93$, $BF_{10} = 2.15 \times 10^3$) and a small, but reliable bias in the *two-way* response distribution away from the first similar probe (**Figure S1B**; M = -1.45°, 95% CI [-2.53, -0.36°], $t(31) = -2.73$, $p = 0.011$, *Cohen's d* = -0.48, $BF_{10} = 4.24$). The absolute magnitude of the bias observed in the *one-way condition* was larger than that in the *two-way condition* ($\Delta M = 3.27°$, 95% CI [1.24, 5.31°], $t(31) = 3.28$, $p = 0.003$, *Cohen's d* = 0.58, $BF_{10} = 1.41 \times 10$). Thus, we again conclude that perceptual comparisons induced reliable biases in memories of real-world objects.

2 Experiment S1: Report biases following perceptual comparisons in a 2AFC paradigm

In Experiment 1, we found that participants' memory reports of colored real-world objects were systematically attracted towards the colors of probe objects that were endorsed as similar during an intervening perceptual comparison. We interpret these report biases as evidence for bona fide changes in the target representation that were caused by the perceptual comparison, either at the time of the comparison or when the target object was accessed again during the memory report. However, findings in Experiment 1 may also be explained by a response strategy in which participants fine-tuned their reports of the target object towards the features of the similar probe object on the color wheel in order to optimize their performance or to communicate their understanding of probe similarity. Critically, this alternative explanation would suggest that target memories remained intact following perceptual comparisons and that the biases were instead driven by systematicity in how the target was reported. Thus, it is unclear whether the biases observed in Experiments 1-2 were driven by the perceptual comparison or a trivial response strategy that was employed during the continuous report of the target.

To adjudicate between these competing explanations, we conducted a control experiment (Experiment S1) where we replaced the continuous report at the end of the trial with a two-alternative forced-choice (2AFC) where participants selected between the correct target and a foil of the target (**Figure S2**). This approach was motivated by previous studies of VWM biases that have used 2AFC paradigms to address response strategies (e.g., Chunharas et al., 2022). The purpose of the 2AFC report is to discourage response strategies that incorporate non-target information by removing the continuous wheel and simply re-presenting the correct target during the memory report. In doing so, participants are unable to use the similar probe as a scaffold to fine-tune their response on the color wheel and their understanding of the task is most straightforwardly communicated by selecting the correct target.

To determine the presence of any true memory biases with the 2AFC report, the foil of the target was sampled slightly towards or away from the similar probe that was endorsed during the perceptual comparison. If the target memory is indeed biased towards the similar probe following perceptual comparisons, then participants should be more likely to select a foil of the target that is sampled towards the similar probe following a perceptual comparison than when a comparison is not made. A memory bias towards the similar probe may also increase the likelihood that participants select the correct target when the foil is sampled away from the similar probe. However, because a memory bias towards the similar probe always makes the target and foil different from memory in this condition, this inverse trend in the 2AFC report may be less reliable.

Method

Participants

Data were collected from 34 undergraduate students in accordance with the procedures approved by the Research Ethics Board at the University of Toronto. All volunteers had normal or corrected-to-normal visual acuity and did not report color-blindness. Two participants' data were removed because of not following task instructions (one) and below chance accuracy in the baseline condition (<50%, one), resulting in 32 participants' data being subjected to analysis (24 female, mean age = 20.5 years old).

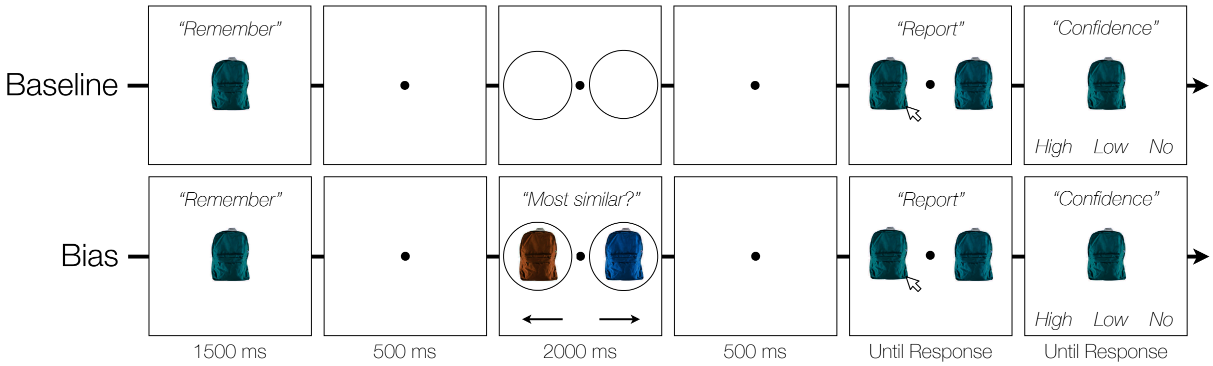

Figure S2. *2AFC Experiment Method*

The trial procedure for the baseline and bias conditions. During the 2AFC report, participants selected between the correct target (*left*) and a foil of the target (*right*) that was sampled 6° towards or away from the similar probe in the underlying feature space. In the figure, the foil is sampled 6° towards the similar probe.

Stimuli and Apparatus

Identical to Experiment 1.

Procedure

Participants performed six blocks of 40 pseudorandomized trials (**Figure S2**). Each trial began with a target object (5.3° x 5.3°) presented at the center of the screen for 1500 ms, which participants were instructed to remember as precisely as possible. Target object colors were randomly sampled from the circular color space (see *Stimuli and Apparatus* in Experiment 1). Target object presentation was followed by a 3000-ms maintenance interval. At the completion of the maintenance interval, two copies of the target object (5.3° x 5.3° each) were presented simultaneously on either side of the screen (5.3° from center). One copy was identical to the target object presented at the start of the trial while the color of the foil copy was slightly distorted (see below). Participants were instructed to select the copy that exactly matched the target object by clicking on it with the mouse cursor. After selection, the selected copy was then displayed by itself at the center of the screen and participants indicated how confident they were that they selected the correct target by pressing one of three keyboard buttons (*high confidence, low confidence, no memory*). The accuracy of the memory report was emphasized and was therefore reported without an imposed time limit.

In half of the trials, participants completed a perceptual comparison during the maintenance interval. 500 ms after the offset of the target object, two copies of the target object (5.3° x 5.3° each) were presented simultaneously as probes on either side of the screen (5.3° from center) inside circular placeholders. The circular placeholders were presented in the baseline and bias conditions to differentiate the perceptual comparison from the memory report at the end of the trial. The color of the similar probe was randomly sampled ±16 - 45° away from the target object. To determine the color of the dissimilar probe, we rotated the sampling window 180° (±196 - 225°) and randomly sampled again. Participants were instructed to indicate which of the two probes was most similar to the original target object by pressing the left or right arrow key on the keyboard. The probes remained on the screen for 2000 ms regardless of the report and were followed by another 500-ms blank delay before the memory report.

The color of the foil that was presented during the memory report was sampled 6° away from the color of the target on every trial and in the direction of the similar probe on half of the trials. Since there was no perceptual comparison performed in the baseline condition, we assigned the offset direction of the foil pseudorandomly in the baseline condition to match the number of trials in the bias condition. We chose the 6° offset value based on two factors. First, 6° was only slightly larger than the average bias magnitude observed in Experiment 1 (=4.72°). Second, 6° rendered the correct target and foil to be very perceptually similar (see **Figure S2** for an example) while still being sufficiently discriminable to allow for above-chance performance (see, e.g., Chunharas et al., 2022 and the *Participants* section above). This high degree of perceptual similarity decreased the likelihood that one of the copies during the memory report was clearly more alike the similar probe than the other.

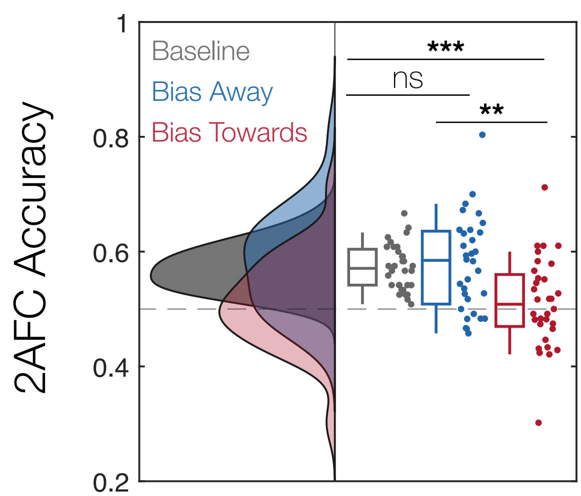

Figure S3. *2AFC Experiment Results*

Boxplots of 2AFC memory report accuracy in the baseline and bias conditions. The bias condition is shown separated by foil sampling conditions. The horizontal line inside each boxplot indicates the median 2AFC accuracy across participants. Colored dots to the right of each boxplot indicate the 2AFC accuracy for a given participant with corresponding density distributions shown on the left-hand panel of the figure. ** $p < 0.01$, *** $p < 0.001$, *ns* = not significant.

Analyses

Identical to Experiment 1, we filtered the data to only include trials where the participant correctly selected the similar probe during the perceptual comparison (M = 97% of trials in the experimental condition). Unlike Experiment 1, the number of high-confidence trials was relatively low and varied considerably between participants (baseline: M = 49.4%, SD = 24.1%; bias: M = 46.5%, SD = 23.4%). This confirmed that our sampling of the foil made the memory report subjectively difficult as intended. To maximize power, we report our results across all trials.

Results & Discussion

To assess the interfering effect of the perceptual comparison, we first compared the accuracy in the 2AFC memory report between the baseline and bias conditions while collapsing across foil sampling conditions. In doing so, we found that participants selected the correct copy more often in the baseline condition than the bias condition, $t(31) = 2.82$, $p = 0.008$, *Cohen's d* = 0.50, BF_{10}

= 5.16. This confirmed that performing a perceptual comparison during maintenance made it more difficult to distinguish between the correct target and the distorted foil.

We then separated trials in the bias condition based on the offset direction of the foil to see if lower performance in the bias condition was linked to the systematic sampling of the foil. If perceptual comparisons result in bona fide attraction biases that are not specific to a continuous report strategy, we should expect that participants selected the incorrect foil more often in the bias condition when it was sampled towards the similar probe than away from it. Consistent with this prediction, we found that 2AFC accuracy in the bias condition was lower when the foil was sampled towards the similar probe than when it was sampled away from it (**Figure S3**, $t(31) = 2.94$, $p = 0.006$, *Cohen's d* = 0.52, $BF_{10} = 6.66$). Importantly, we found no difference in 2AFC accuracy between the foil sampling conditions in the baseline condition, where the offset direction of the foil was assigned pseudorandomly, $t(31) = 0.06$, $p = 0.951$, *Cohen's d* = 0.01, $BF_{01} = 5.29$. This null difference in 2AFC accuracy between the foil sampling conditions in the baseline condition confirmed that the significant differences in the bias condition were unlikely to be explained by chance alone.

Yet, one could argue that the difference in 2AFC accuracy between the foil sampling conditions in the bias condition was still due to a response strategy. For example, it may have been the case that participants could clearly discern which copy was more alike the similar probe on some trials and selected it to try and optimize their performance. To address this concern, we compared 2AFC accuracy in each foil sampling condition to the overall accuracy in the baseline condition. If the difference in 2AFC accuracy between foil sampling conditions in the bias condition was the result of a response strategy that was implemented only on select trials, we should expect that this response bias would not be reliable enough to produce a difference from baseline. However, as shown in **Figure S3**, 2AFC accuracy was lower in the bias condition than the baseline condition when the foil was sampled towards the similar probe ($t(31) = -4.59$, $p < 0.001$, *Cohen's d* = -0.81, $BF_{10} = 3.53 \times 10^2$), suggesting a reliable bias across trials. Interestingly, we also found that 2AFC accuracy was comparable between the baseline and bias conditions when the foil was sampled away from the similar probe in the bias condition (**Figure S3**, $t(31) = 0.36$, $p = 0.720$, *Cohen's d* = 0.06, $BF_{01} = 4.98$), confirming that participants were not just selecting the option in the 2AFC report that looked more like the similar probe on every trial. Thus, we conclude that report biases observed in Experiments 1 and 2 were unlikely to be explained by a response strategy that was specific to the continuous report and was more likely indicative of a bona fide bias caused by the perceptual comparison.

3 Memory bias in Experiment 2 for all confidence levels

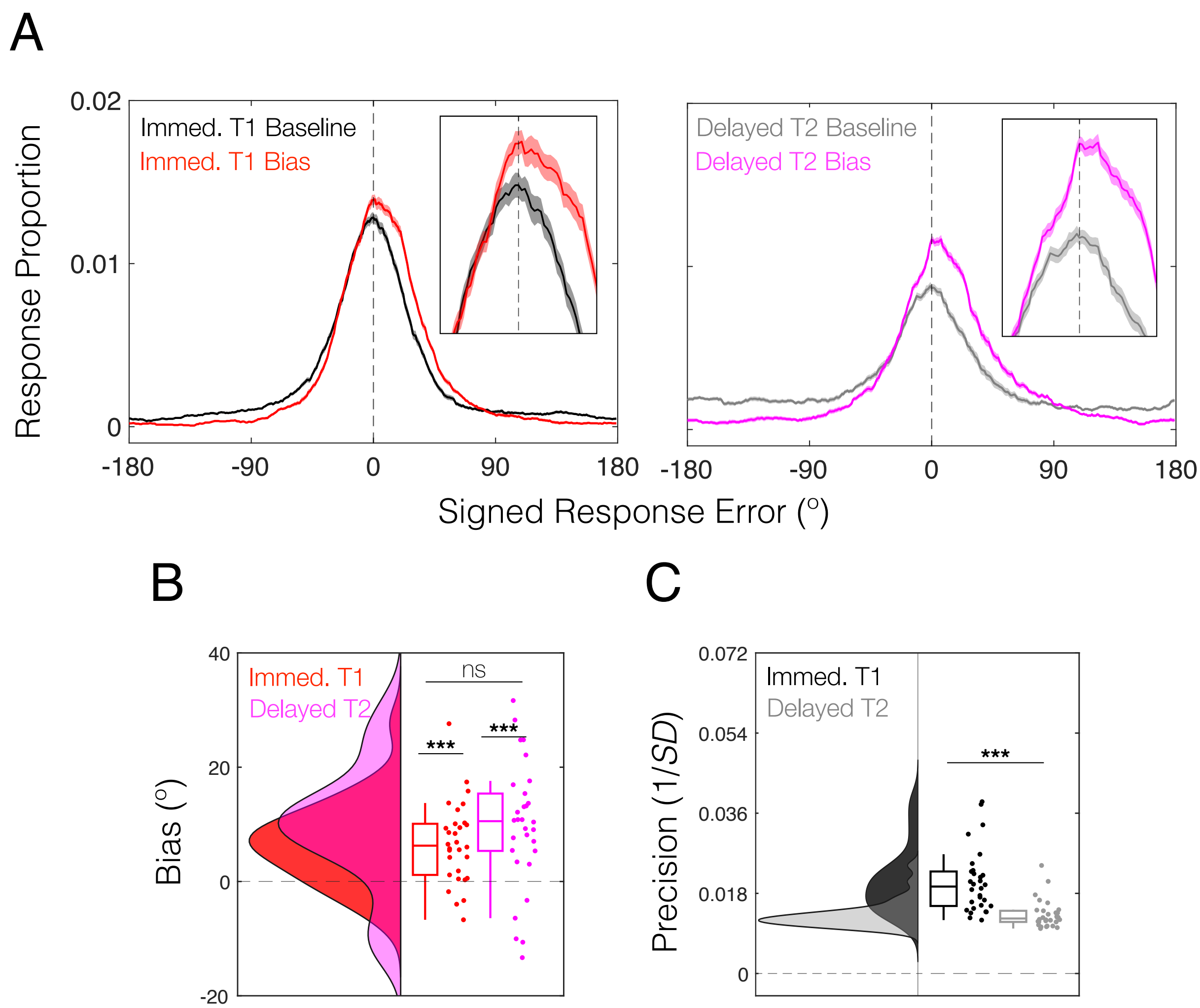

Figure S4. *SIMB Present in Long-Term Memories with All Trials Included.*

(**A**) Signed response distributions for set of encoded objects (i.e., immediate, delayed) at each memory test (i.e., Test 1, Test 2) across all trials. For demonstration purposes, we plotted the proportion of responses for a given signed offset value by calculating the mean response proportion across a 30° window centered at the offset value. Positive offsets indicate memory bias towards the similar 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 dashed line indicates the location of the target. (**B**) Boxplots of the mean signed response error (i.e., bias) in each set of encoded objects. Positive values indicate memory bias towards the similar probe. (**C**) Boxplots of the inverse standard deviation of response errors (i.e., precision) in each set of encoded objects that were recalled in the baseline condition. Larger values indicate better precision. The horizontal line inside every boxplot indicates the median value across participants. Colored dots to the right of each boxplot indicate the value for a given participant with corresponding density distributions shown on the left-hand panel of the figure. T1 = Memory Test 1, T2 = Memory Test 2, SD = standard deviation, ns = not significant, *** $p < 0.001$

Signed response distributions for Immediate Set objects at Test 1 and Delayed Set objects at Test 2 are shown in **Figure S4A**. Across confidence levels, memory biases in each object set were

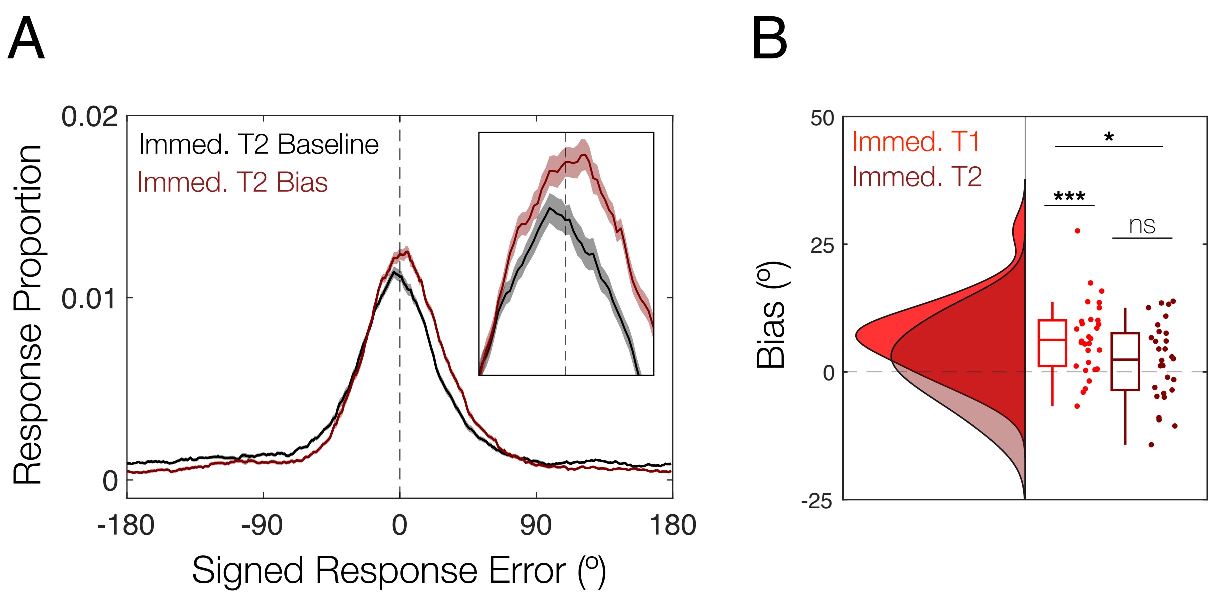

Figure S5. *SIMB Persistence Across 24-hour Delay with All Trials Included.*

(**A**) Signed response distributions at Memory Test 2 for Immediate Set Objects across all trials. For demonstration purposes, we plotted the proportion of responses for a given signed offset value by calculating the mean response proportion across a 30° window centered at the offset value. Positive offsets indicate memory bias towards the similar probe that was presented at Memory Test 1. 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 dashed line indicates the location of the target centered across trials. (**B**) Boxplots of the mean signed response error (i.e., bias) at Memory Test 1 and 2 for Immediate Set Objects across all trials. Positive values indicate memory bias towards the similar probe that was presented at Memory Test 1. The horizontal line inside each boxplot indicates the median bias across participants. Colored dots to the right of each boxplot indicate the bias for a given participant with corresponding density distributions shown on the left-hand panel of the figure. T1 = Memory Test 1, T2 = Memory Test 2, ns = not significant, * $p < 0.05$, *** $p < 0.001$

significant (**Figure S4B;** Test 1: $M = 6.66°$, 95% CI [4.00, 9.33°], $t(29) = 5.11$, $p < 0.001$, *Cohen's d* = 0.93, $BF_{10} = 1.20 \times 10^3$; Test 2: $M = 9.70°$, 95% CI [5.58, 13.82°], $t(29) = 4.82$, $p < 0.001$, *Cohen's d* = 0.88, $BF_{10} = 5.68 \times 10^2$). However, memory biases in Delayed Set objects at Test 2 were only numerically larger than those in Immediate Set objects at Test 1 ($\Delta M = 3.04°$, 95% CI [-1.53, 7.61°], $t(29) = 1.36$, $p = 0.184$, *Cohen's d* = 0.25, $BF_{01} = 2.23$). When we compared the persistence of memory biases from Test 1 to Test 2, we found that biases significantly decreased ($\Delta M = 4.61°$, 95% CI [0.92, 8.30°], $t(29) = 2.56$, $p = 0.016$, *Cohen's d* = 0.47, $BF_{10} = 3.01$) and were no longer significant at Test 2 (**Figure S5**; $M = 2.05°$, 95% CI [-0.77, 4.88°], $t(29) = 1.49$, $p = 0.148$, *Cohen's d* = 0.27, $BF_{01} = 1.91$). Thus, when low-confidence reports and guesses were included in the analysis, results observed within high-confidence reports were partially eliminated.

These findings are consistent with our prior prediction (see *Analysis* sections in main manuscript) stating that low-confidence memory reports may increase the response variability in the dataset and make it difficult to detect reliable biases. Low confidence is typically indicative of a low-quality memory or one that has been forgotten entirely. In these situations, individuals may rely on information outside the target memory to complete the ongoing task. Thus, the present results do not contradict the robust and persistent biases observed in high-confidence reports, where individuals consistently relied on a true memory signal. Instead, one may argue that the patterns that were still preserved in some of the present comparisons (including Experiment 1, **Figure S1**) underscore the large effect sizes associated with memory biases. Thus, we still conclude that perceptual comparisons induce memory biases in long-term memory representations, and that these biases persist when individuals are highly confident in the accuracy of their report.

References

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*. Advance online publication. https://doi.org/10.1037/xge0001191