# Distractors induce space-specific neural biases in visual working memory

**Authors & Affiliations:**

Deepak V Raya[1], Sanchit Gupta[1], and Devarajan Sridharan[1,2,*]

[1]Centre for Neuroscience, Indian Institute of Science, Bangalore 560012, India

[2]Computer Science and Automation, Indian Institute of Science, Bangalore 560012, India

***Correspondence to:**

Devarajan Sridharan,

GS Bhat Chair Associate Professor in Digital Health,

DST SwarnaJayanti Fellow,

Centre for Neuroscience,

Indian Institute of Science,

C. V. Raman Avenue,

Bangalore 560012, India

Phone: +91 802 293 3434

E-mail: sridhar@iisc.ac.in

## Abstract

Information held in working memory is remarkably resilient to interference. Yet, distractors containing task-relevant features impact working memory systematically. The neural basis of this interference is poorly understood. Combining behavior with multivariate decoding of human electroencephalography recordings, we show that distractors induce space-specific biases in mnemonic representations. Participants performed a visual working memory task involving grating orientation estimation, with cued and uncued memoranda appearing on opposite visual hemifields. A high-contrast grating distractor, flashed briefly during the delay epoch, biased both behavioral reports and neural representations. Mnemonic representations were biased either towards the distractor's orientation (attractive) or away from it (repulsive) depending, respectively, on whether the distractor appeared in the same hemifield as the corresponding memorandum, or opposite to it. On the other hand, the magnitude of the bias was much stronger for the uncued than the cued, memorandum, and depended critically on the distractor's timing, encoding strength and gating efficacy. A two-tier ring attractor model incorporating cross-hemifield competitive inhibition, parsimoniously explained each of these experimental observations. Overall, our results show that prioritization in working memory shields mnemonic representations from distractor interference, and provide a testable computational hypothesis for distractor effects in visual working memory.



**Keywords**: working memory; distractor interference; attractive and repulsive biases; mnemonic representation; multivariate decoding; ring attractor model.

## Introduction

Working memory (WM) is a remarkable cognitive ability that enables us to transiently hold and manipulate information that is no longer present in the environment. Mnemonic information in WM must be maintained in the absence of external sensory drive, and it is increasingly evident that early sensory areas are actively involved in such maintenance[1–5]. As a result, mnemonic information may be susceptible to interference by new sensory inputs, such as salient distractors[5–7]. Yet, a significant body of literature suggests that information in WM is surprisingly resilient to distraction[5,6,8,9]. For example, behavioral performance in a WM task involving oriented gratings was robust, despite multiple task-irrelevant (face or place) distractors being presented during the delay period[6]. Similar results have been observed in many recent studies of visual WM[5,10].

Despite this robustness, recent evidence suggests that distractors may interfere subtly, albeit systematically, with the content of WM[5,8,10,11]. Some distractors are more potent than others[8,9,11–16] and, in particular, certain distractors reliably bias the content of working memory. The extent to which a distractor's features match task-relevant features of the memorandum determines the strength of its bias, with greater feature similarity yielding greater bias in WM[8,15]. Typically, such biases are "attractive" in that the memorandum's report is biased towards the distractor's features[8,10,11,17]. For instance, viewing a task-irrelevant orientation during the WM delay period biases the remembered target's orientation towards the distractor's orientation[10,17]. Similarly, spatial distractors can bias the remembered location towards the distractor's location[11]. Such distractor biases have been reported even for complex features. For example, remembered face representations are more strongly biased by the presentation of face image distractors than by scene image distractors[15].

Identifying mechanisms that mediate distractor interference, including biases, in neural WM representations is a topic of active research[5,8–11]. Previous studies measured brain responses during the WM delay epoch, and have found evidence for a distributed representation of mnemonic information in diverse brain regions[18], including sensory (visual) cortex[1–5], posterior parietal cortex (PPC, [10,19,20]), the intraparietal sulcus (IPS, [6,10,21]), as well as in the prefrontal cortex (PFC, [7,21,22]). Converging evidence from these studies suggests that while mnemonic information in sensory cortex may be susceptible to distractor interference and bias, information in higher cortical areas, including PPC[6,10] and PFC[9,23], is much more robust to such interference. For example, when a distractor grating with a random orientation was presented during the delay period of a working memory task, it produced biases in neural representations in the early visual cortex (V1-V3), but not in the IPS[10]. A related study showed that the distractor's orientation affected neural representations in early visual cortex (V1-V4), but higher cortical regions, including the IPS and lateral occipital cortex, were unaffected[5]. In fact, such stability of representations in higher cortical areas is hypothesized to mediate the remarkable resilience of WM to sensory distractors [8,24].

Despite this emerging literature, a key question remains unanswered: how does a sensory distractor bias neural representations in working memory, depending on its location and timing? Previous studies on distractor effects overwhelmingly employed simple WM task designs: for example, most tasks involved a single, central distractor presented at fixed times during the delay epoch[5,8–12,14–16,25]. Moreover, many more previous studies investigating mnemonic bias employed functional magnetic resonance imaging (fMRI)[5,10,11] to study WM biases in the brain. The poor temporal resolution of fMRI precludes addressing fundamental questions about rapid, distractor timing effects: Because the memorandum and distractor occur within a few hundred milliseconds of each other, their evoked blood oxygenation level dependent (BOLD) responses evolve over the timescale of several seconds and are challenging

to disentangle. No previous study, to our knowledge, has explored distractor-induced bias effects by varying the distractor's location and its timing relative to the memorandum, systematically.

Here, we addressed these challenges with multivariate stimulus decoding of high-density human electroencephalography (EEG) recordings, which enabled measuring distractor's neural effects with millisecond precision. We characterized how distractors induce mnemonic biases across space and time in participants (n=24) performing a continuous response, visual WM task (Fig. 1A). With state-of-the-art neural decoding we show that a task-irrelevant distractor grating presented during the delay period produces either an attractive or a repulsive bias on mnemonic neural representations depending, respectively, on whether the distractor was presented on the same or the opposite hemifield as the corresponding memorandum. In addition, prioritization in working memory strongly shields mnemonic representations from distractor interference. Finally, we synthesize these findings with a two-tier ring attractor model[26,27], that provides a testable mechanistic explanation for these distractor-induced biases in visual working memory.

## Results

### Distractors induce a space-specific bias in visual working memory

We tested n=24 human participants on a visual WM task involving grating orientation report (Fig. 1A, Methods). Following fixation, two randomly oriented gratings ("memory array") were flashed briefly (100 ms) one in each visual hemifield. A brief interval (500 ms) after grating offset, a spatial retro-cue (filled white semi-circle) appeared, which indicated the side of the erstwhile grating that would likely be probed for response (cue validity: 70%). A fixed epoch

(2450 ms) after the memory array offset, a probe appeared (central filled circle) whose color (white or black) indicated the grating (cued or uncued, respectively) whose orientation the participants must reproduce (Methods).

To study whether space-specific biases were induced by distractors in visual WM, a singleton, distractor grating – randomly oriented but otherwise identical to the stimuli in the memory array (Methods) – was flashed briefly (100 ms) at a random time (500-1250 ms, exponentially distributed) after retro-cue offset. The distractor was presented in 80% of the trials – interleaved pseudorandomly with 20% “no distractor” trials – and appeared either on the left or the right hemifield with equal probability. We refer to the retro-cued hemifield, the grating at this location or its representation in WM as the “cued” hemifield, grating, or representation, respectively; their counterparts in the opposite hemifield are referred to as the “uncued” hemifield, grating or representation, respectively.

First, we tested if participants were indeed utilizing the retro-cue for selection in visual WM, including only trials in which no distractor appeared (20% of all trials). The error in orientation reports, as measured by the circular standard deviation (CSD) of the orientation report error, was significantly lower on cued trials compared to uncued trials ($CSD_{Cued}$=23.4°±1.6° [mean±s.e.m.]; $CSD_{Uncued}$=35.2°±2.2°, $p<0.001$, Wilcoxon signed rank test, Bayes Factor ($BF_{10}$) $>10^2$; n=24 participants) (Fig. 1B), confirming robust selection of the retro-cued representation in visual WM.

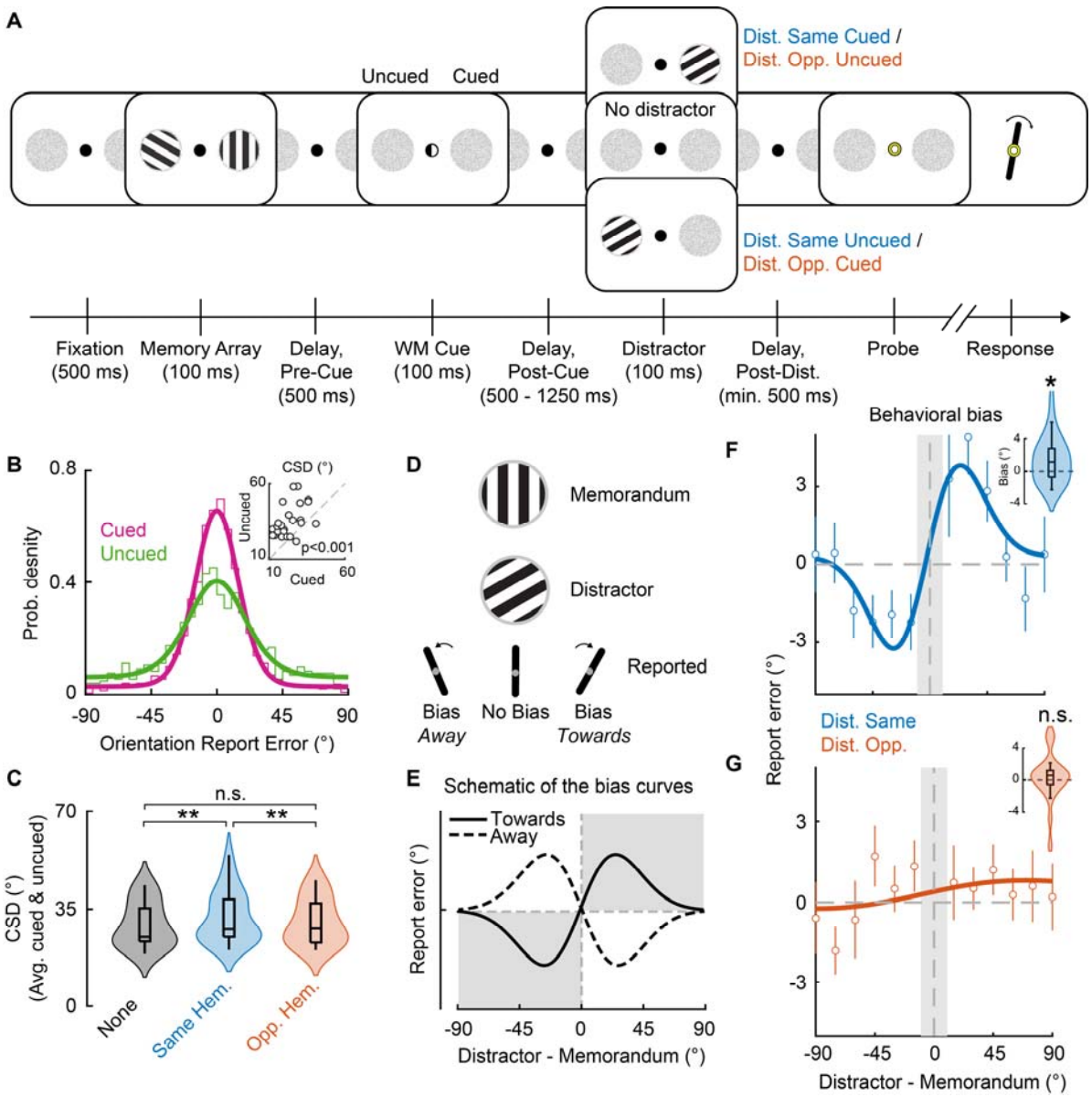



**Figure 1. Distractor biases behavioral estimates of memorandum orientation.**

**A.** Visual working memory task with a distractor. Two randomly oriented gratings were flashed briefly one in each hemifield. A retro-cue (white semicircle) indicated the location of the erstwhile grating that would likely be probed for response (70% validity). Following a variable delay (post-cue), on 80% of the trials, a high contrast distractor grating – identical in size and spatial frequency as the memoranda – was flashed either on the right or left hemifield (upper and lower rows); on the remaining trials (interleaved) no distractor appeared (middle row). At the

end of the trial, a central response probe (white or black) indicated the location of the grating (cued or uncued, respectively) to be reproduced.

**B.** Distribution (thin histogram) of the orientation report errors for the cued (magenta) and uncued (green) trials (n=24, data pooled across participants). Thick lines: von Mises fits. (*Inset*) Circular standard deviation (CSD) of the report errors for “no distractor” trials, for cued (x-axis) versus uncued (y-axis) grating reports. Dashed line: line of equality. Data points: individual participants.

**C.** CSD for no distractor trials (black), when the distractor appeared in the same (blue), or opposite (orange) hemifield as the probed memorandum. Violin plots: kernel density estimates; center line: median; box limits: upper and lower quartiles; whiskers: 1.5x interquartile range.

**D.** The distractor (lower grating) may bias the memorandum’s (upper grating) report away from itself (left bar), towards itself (right bar), or not at all (middle bar).

**E.** Schematic of bias curves. Error in the probed memorandum’s orientation (signed) (y-axis) as a function of its orientation relative to the distractor (x-axis). Positive and negative values (both axes): clockwise and counter-clockwise relative orientations, respectively. Solid and dashed curves: biases toward, or away from, the distractor’s orientation, respectively. Gray shading: strength of bias (Methods).

**F.** Average behavioral bias (n=24) when the distractor appeared in the same hemifield as the probed grating (data pooled across cued and uncued memoranda). Solid line: Derivative of Gaussian fit (Methods). Error bars: s.e.m. Gray shaded rectangle: untested relative orientation values (±10°). (*Inset, top right*) Violin plot: bias strengths across participants. Other conventions are the same as in panel C.

**G.** Same as in panel F, but for trials in which the distractor appeared in the opposite hemifield as the probed grating. Other conventions are the same as in panel F. (All panels) *: $p<0.05$; **: $p<0.01$; n.s.: not significant.

Next, we analyzed the effect of the distractor on the cued and uncued gratings' orientation reports including all trials (80%) in which the distractor was presented. We tested for spatially lateralized effects; specifically, we hypothesized that the precision of orientation reports would be different depending on whether the distractor was presented in the same ("distractor-same" condition) or on the opposite hemifield ("distractor-opposite" condition) as the probed stimulus. We found robust evidence confirming this hypothesis: a 2-way ANOVA with the CSD of orientation errors as the dependent variable revealed a significant main effect of distractor location ($F_{(1,23)}=11.06$, $p<0.001$), but no significant interaction effect between cueing and distractor location ($F_{(1,23)}=0.48$, $p=0.625$). Post-hoc analyses revealed a significant increase in error for distractor-same compared to no-distractor trials (Fig. 1C, $CSD_{NoDist}=29.3°\pm1.7°$, $CSD_{DistSame}=32.0°\pm1.9°$, $p=0.002$, BF=40.34, signed rank), but not for distractor-opposite compared to no-distractor trials (Fig. 1C, $CSD_{NoDist}=29.3°\pm1.7°$, $CSD_{DistOpp}=30.3°\pm1.7°$, $p=0.103$, BF=1.16, signed rank); data pooled across cued and uncued locations). In other words, the distractor affected the precision of orientation reports for the memorandum when it occurred, subsequently, in the same hemifield but not when it occurred in the opposite hemifield.

It is possible that the increase in orientation report error was merely a consequence of non-specific interference in WM maintenance by the distractor – a salient, flashed stimulus. Alternatively, it is possible that the increase in error arose from a specific bias in WM content induced by the distractor's orientation. To disambiguate these hypotheses, we tested whether the behavioral orientation report of the probed memorandum was biased by the distractor's orientation, by plotting the signed orientation report error for the probed grating as a function of

the difference between the distractor and the probed gratings' orientation; bias was quantified with the signed area-under-the-curve metric[28] (Fig. 1E) (Methods).

Orientation reports for the probed memorandum were systematically biased towards the distractor's orientation when the distractor appeared in the same hemifield as the former (Fig. 1F) ($Bias_{DistSame}$=1.35°±0.58°, p=0.01, $BF_{+0}$=4.06; data pooled across cued and uncued gratings). By contrast, no significant bias effects occurred when the distractor appeared in the opposite hemifield as the probed memorandum (Fig. 1G) ($Bias_{DistOpp}$=0.34°±0.6°, p=0.7, $BF_{-0}$=0.14). Next, we analyzed these biases separately for the cued and uncued grating, based on subsets of trials in which each was probed for response. The overall bias effects were replicated for the cued grating (SI Fig. S1 A-B) (Cued: $Bias_{DistSame}$=0.94°±0.35°, p=0.004, $BF_{+0}$=7.50; $Bias_{DistOpp}$=-0.3°±0.3°, p=0.17, $BF_{-0}$=0.54). Qualitatively similar trends were apparent for the uncued grating also, particularly for the "same side" distractor trials (SI Fig. S1, C-D) but these trends did not reach significance (Uncued: $Bias_{DistSame}$=1.76°±1.15°, p=0.07, $BF_{+0}$=1.12; $Bias_{DistOpp}$=0.98°±1.07°, p=0.8, $BF_{-0}$=0.11), possibly because of the comparatively fewer trials (30%) in which uncued gratings were probed.

In sum, the distractor induced an attractive bias for orientation reports when it occurred in the same hemifield as the probed memorandum, but not when it occurred in the opposite hemifield. This space-specific bias induced by the distractor suggests that maintenance in working memory is mediated, at least in part, by a spatially lateralized, hemisphere-specific process. To pinpoint mechanisms underlying this bias, and to identify its neural origins, we analyzed neural representations of the memory array and distractor with multivariate stimulus decoding of electrophysiological recordings.

**Distractors induce antagonistic biases in mnemonic neural representations**

To better understand the neural basis of the bias induced by the distractor on the memoranda, we computed the bias in the underlying neural representations. For this, first, we performed trial-wise decoding of the memorandum and distractor orientation from time-resolved EEG activity using a multivariate decoder based on the pair-wise Mahalanobis distance[29,30]. The orientation of both memory array gratings could be decoded robustly from the (respective) contralateral electrodes shortly following their onset (Fig. 2A, B) ($p<0.001$, cluster-forming threshold $p<0.05$; n=23 participants for whom EEG data was available, see Methods).

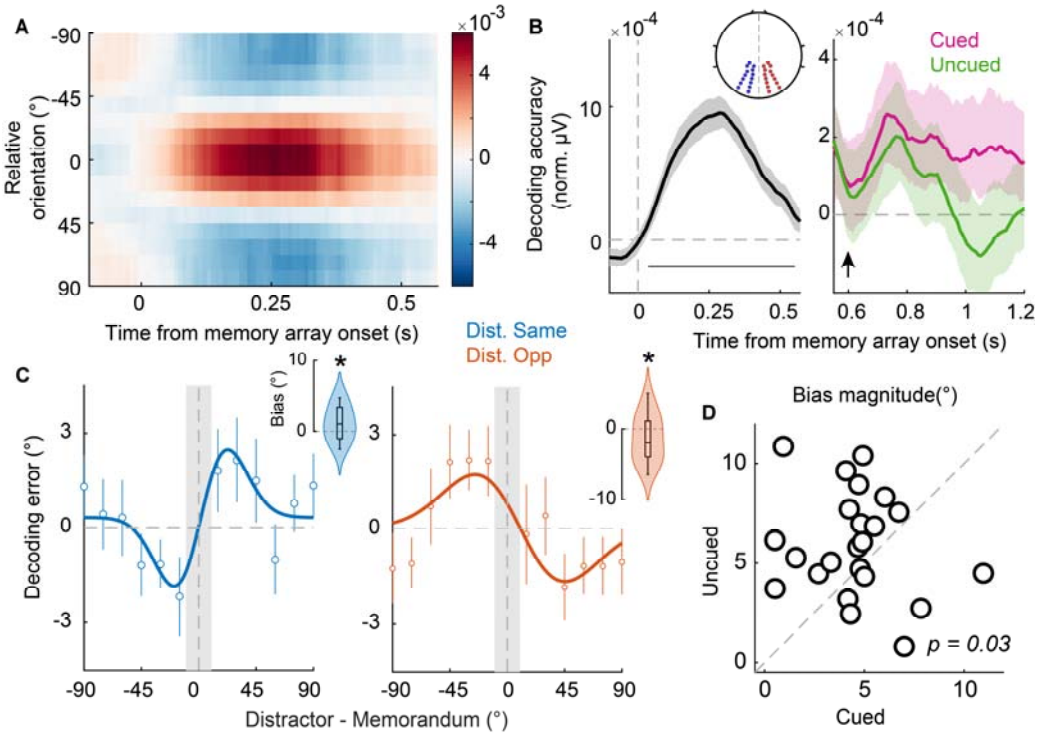



**Figure 2. Distractor biases mnemonic neural representations.**

**A.** Orientation tuning map for memory array decoding for each time point relative to memory array onset. The tuning is computed as the similarity (sign reversed, mean-centered Mahalanobis distance) between the neural activity of the test trial and reference trials for stimuli at different relative orientations (y-axis). (All panels) Unless indicated otherwise, data averaged across participants (n=23) and for the left and the right gratings.

**B.** (*Left*) Decoding accuracy (dot product of tuning curve with cosine, normalized uV, Methods) for each time point locked to the memory array onset (dashed vertical bar). Horizontal black bar: significant decoding epochs (cluster-based permutation test, Methods). Shading: s.e.m. Profiles were smoothed with an 80 ms rectangular moving window for clarity of visualization. *(Inset)* Posterior electrodes used for the decoding analysis, contralateral to the respective grating. (*Right*) Same as in the left panel but showing decoding accuracy separately for cued (magenta) and uncued (green) gratings following cue onset (arrowhead).

**C.** Neural bias curves, following the same format as in Figure 1E, but showing the decoding error in the neural representation of the memorandum for trials in which the distractor appeared in the same hemifield (*left*) or opposite hemifield (*right*) relative to the respective memorandum; data pooled across cued and uncued memoranda. Error bars: s.e.m. (*Inset*) Distribution of the neural bias across participants. *: $p<0.05$; **: $p<0.01$; n.s.: not significant. Other conventions are the same as in Figures 1F-G.

**D.** Magnitude (absolute value) of bias strength for the cued (x-axis) versus the uncued (y-axis) memoranda. Data averaged across distractor locations. Circles: individual participants. Dashed line: line of equality.

Next, we analyzed neural biases in the decoded memory array orientation during the maintenance epoch following the appearance of the distractor (Fig. 1A, Delay, Post-Dist.). To avoid distractor's evoked response affecting the memorandum's decoded orientation, we analyzed data in a ±150 ms window centered at 350 ms following distractor onset, with trials pooled across the cued and uncued memoranda. The decoded orientations of the memoranda were biased towards the distractor's orientation (attractive bias) when the distractor was presented in the same hemifield as the respective memorandum (Fig. 2C, left) ($Bias_{DistSame}$=1.11°±0.48°, p=0.010, $BF_{+0}$=3.83). By contrast, decoded orientations of the memoranda were biased away from the distractor's orientation (repulsive bias) when the

distractor was presented in the opposite hemifield as the respective memorandum (Fig. 2C, right) ($Bias_{DistOpp}$= -1.31°±0.68°, p=0.03, $BF_{-0}$=2.00). Analyzing the cued and uncued memoranda separately, we found that while the cued memorandum manifested little or no evidence of bias (SI Fig. S2 A-B) ($Bias_{DistSame}$: p=0.56, $BF_{+0}$=0.2; $Bias_{DistOpp}$: p=0.35, $BF_{-0}$=0.31), the uncued representation was strongly biased by the distractor in both conditions (SI Fig. S2 C-D) ($Bias_{DistSame}$: p=0.004, $BF_{+0}$=9.46; $Bias_{DistOpp}$: p=0.020, $BF_{-0}$=2.40). A 2-way ANOVA revealed a significant main effect of cueing on signed neural bias magnitude ($F_{(1,22)}$=7.4, p=0.01); post-hoc tests revealed that absolute magnitude of the distractor-induced neural bias, was greater for the uncued, as compared to the cued, gratings (Fig. 2D) (p=0.03, BF=1.03; signed rank test). In other words, the neural representation of the cued memorandum was more robust to interference by distractor information, as compared to the uncued memorandum, in VWM.

We performed three control analysis to verify the robustness of these findings. First, to examine whether the memorandum could be decoded across the delay period, despite the intervening appearance of the distractor we computed orientation decoding accuracy in two windows: i) an early window 500 ms before the earliest distractor onset (SI Fig. S3A) and ii) a later window 500 ms after the final distractor offset (SI Fig. S3B, left). Memorandum decoding performance was significantly above chance in the early window (p=0.04, BF=1.65), and not significant in the late window (p=0.21, BF=0.44); interestingly, in the late window decoding was significantly above chance for cued (p=0.03), but not the uncued (p=0.43) memoranda (SI Fig. S3B, right). These results confirm that the cued memorandum's neural representation was more robust to interference by the distractor; the degradation of the uncued representation occurred because the bias induced by the distractor on the uncued side likely persisted late into the mnemonic encoding period. Second, to account for the potential "leakage" of the distractor information into memorandum decoding, we trained an orientation decoder for each memorandum at each time sample from -250 ms to 700 ms following distractor onset and tested

if it could predict the distractor's orientation at the corresponding time samples. If the biases we report merely reflect the distractor's decoded neural representation, we should observe significant evidence for cross-generalized decoding between the memorandum and the distractor during the delay period. Controverting this hypothesis, this "cross-decoder" performed at chance with decoding the distractor's orientation regardless of whether it appeared in the same or the opposite hemifield as the memorandum (SI Fig. S3C) ($p>0.6$, cluster-based permutation test, no clusters significantly above chance). Third, we repeated these analysis by computing biases with electrodes from bilateral occipital cortex (SI Fig S3D, inset), identified based on literature on decoding memoranda with EEG[29]; this control analysis was performed to test if the pattern of biases we found arose from choosing contralateral electrodes alone for memorandum decoding (Fig. 2C). Again, we found the same pattern of results as reported earlier, namely, antagonistic biases produced by the distractor depending on where it appeared in space relative to the respective memorandum (SI Fig. S3D) ($Bias_{DistSame}$: $p=0.06$, $BF_{+0}=1.22$; $Bias_{DistOpp}$: $p=0.002$, $BF_{-0}=1.3$).

In sum, the mnemonic neural representation of the memoranda exhibited antagonistic patterns of biases, depending on their location relative to the distractor: A distractor in the same (opposite) hemifield as the erstwhile memorandum produced an attractive (repulsive) bias toward (away from) its orientation. These effects were apparent for the uncued memorandum, but not the cued memorandum, indicating that selection in visual WM protects robustly against distractor interference.

**Distractor timing, encoding strength and input gating control bias magnitude**

To further understand the mechanistic basis of these distractor-induced biases, we investigated specific factors that might affect the strength of these biases in WM. Based on

recent literature[8], we hypothesized that i) the strength of the distractor's encoding following presentation, ii) its timing during the maintenance epoch, and iii) its input gating would modulate distractor-induced bias.

While the first two factors (encoding strength, timing) are readily quantified (see next), there is no well-established metric, to our knowledge, of distractor input gating[8]. Consequently, to quantify this factor, we leveraged our finding that information in the prioritized location (cued side) was shielded more strongly against distractor interference. Therefore, we explored neural signatures associated with the distractor that exhibited marked hemifield asymmetries between the cued and uncued sides. We quantified and compared three neural signatures: i) the strength of distractor encoding, given by its neural decoding accuracy, ii) the bias in the distractor's neural representation induced by the memorandum's orientation on each side and iii) distractor-evoked event-related potential (ERP) components, N1 and P2/P3a, which are known to index the allocation of spatial attention and working memory, respectively[31–33].

The strength of distractor encoding was not different based on whether it appeared on the cued or the uncued side (Fig. 3A) (decoding accuracy: Cued=$9.9\pm2.6\times10^{-4}$, Uncued=$9.8\pm2.60\times10^{-4}$, p=0.98, BF=0.22). Nor was there a bias in the distractor's neural representation induced by the cued or the uncued gratings (Fig. 3B) (Bias: Cued=-0.12±0.65°, p=0.85, BF=0.22; Uncued=-0.66±0.99°, p=0.5, BF=0.26). But when we compared ERP components, although the distractor-evoked contralateral N1 amplitude (Methods) was not different between hemifields (Fig. 3C) (Cued=-0.85±0.22, Uncued=-0.84±0.21, p=0.73, BF=0.22) the contralateral P2/P3a amplitude was significantly higher for the cued than the uncued hemifield (Fig. 3C) (Cued=0.91±0.19, Uncued=0.72±0.18, p=0.015, BF=4.135). Consequently, we quantified P2/P3a amplitude as a potential indicator of distractor input gating (see also Discussion, for related literature).

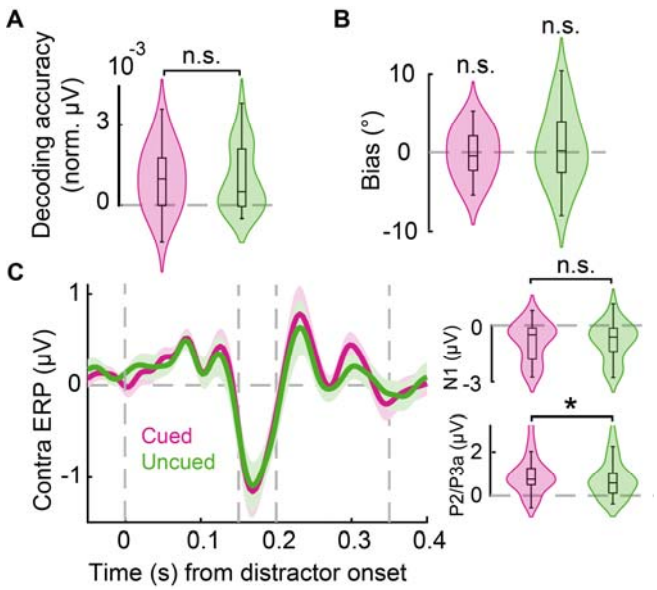



**Figure 3. A putative marker of distractor gating into working memory.**

**A-B.** Distribution of distractor orientation decoding accuracy (**A**) or neural bias in the distractor representation (**B**) when the distractor appeared in the in the cued (magenta) or the uncued (green) hemifields (n=23 participants). Other conventions are the same as in Figure 2C.

**C.** *(Left)* Event-related potentials (ERP) locked to distractor onset (dashed vertical line at 0s) over posterior electrodes contralateral to distractor. Magenta and green traces: ERPs evoked by distractor in the cued and the uncued hemifields, respectively. Solid vertical lines at 0.15s, 0.2s and 0.35s demarcate the boundaries for computing N1 and P2/P3a amplitudes, respectively. Shading: s.e.m. (*Insets*, *right*) Distribution of the N1 (*top*) and P2/P3a (*bottom*) amplitudes evoked by the distractor in the cued (magenta) and uncued (green) hemifields, respectively. Other conventions are the same as in panel A. (All panels) *: $p<0.05$; **: $p<0.01$; n.s.: not significant.

Based on this definition, we tested whether the distractor's timing, encoding strength and input gating would affect the magnitude of distractor-induced bias in behavioral orientation reports. As a first step, we tested and confirmed that these factors were not mutually correlated with each other (SI Results Section on *Distractor timing, strength of encoding and gating are*

*independent factors*). Next, we divided (median split) the trials based on the distractor's latency, its decoding accuracy or the P2/P3a amplitude. We then computed the distractor-induced orientation bias, when the distractor appeared on the same hemifield as the probed memorandum, for both groups of trials. We tested whether the bias for each half of the median split was, individually, significantly different from zero, and also tested for within-participant (pairwise) differences across the median split halves. Moreover, due to the relative paucity of trials in which the uncued grating was probed for behavior, we performed these analyses with only the trials in which the cued grating was probed.

For the first analysis, we divided trials based on the time at which the distractor appeared during the working memory epoch; we term these "early" and "late" distractor trials. Because our task design incorporated a fixed delay between the memory array and response probe presentation, early (late) distractors appeared earlier (later) during the WM maintenance epoch, and were temporally more distal (proximal) to the response probe (Fig. 1A). Distractor-induced orientation bias, when the distractor appeared on the same hemifield as the cued grating, was significantly positive for late distractor trials ("late": behavioral $Bias_{DistSame}$=1.6°±0.4°, p<0.001, $BF_{+0}>10^2$) (Fig. 4A). By contrast, this bias was not significantly different from zero for early distractor trials ("early": behavioral $Bias_{DistSame}$=-0.3°±0.8°, p=0.64, $BF_{+0}$=0.16) (Fig. 4A). Moreover, the bias was significantly stronger for late distractor compared to early distractor trials (p=0.014, BF=3.4).

For the second analysis, we divided trials based on the strength of the distractor's decoded orientation quantified with the cosine similarity metric (Methods); we term these the "weak" and "strong" distractor trials. Distractor-induced orientation bias, when the distractor appeared on the same hemifield as the cued grating, was significantly positive for strong distractor decoding trials ("strong": behavioral $Bias_{DistSame}$=1.79°±0.74°, p=0.006, $BF_{+0}$=4.70) (Fig. 4B). By contrast, this bias was not significantly different from zero for weak distractor

decoding trials ("weak": behavioral $Bias_{DistSame}$=0.96°±0.75°, p=0.11, $BF_{+0}$=0.80) (Fig. 4B). In this case, we did not observe a significant difference between the strong and weak encoding trials (p=0.217, BF=0.43).

For the final analysis, we divided trials based on the amplitude of the distractor-evoked contralateral P2/P3a component, for trials in which distractor appeared on the same hemifield as the cued grating. Distractor-induced orientation bias was not significantly different from zero for high P2/P3a-amplitude trials ("strong": behavioral $Bias_{DistSame}$=0.63°±0.68°, p=0.20, $BF_{+0}$=0.52) (Fig. 4C). By contrast, this bias was significant, and attractive, for weak P2/P3a amplitude trials ("weak": behavioral $Bias_{DistSame}$=2.4°±0.66°, p<0.001, $BF_{+0}$=47.03) (Fig. 4C). Here, we observed a marginally significant difference between the high and low P2/P3a amplitude trials (p=0.030, BF=1.90). In all three cases, despite robust effects for the individual splits, testing for pairwise differences yielded weaker effects; in this case, bias estimates were likely rendered noisier with only half of the number of trials within each split.

Overall, these results show that distractors that appeared later during working memory maintenance, as well as those that were encoded more robustly, produced a stronger bias in orientation reports of the cued memorandum. Moreover, the effectiveness of distractor input gating – as quantified operationally by the distractor-evoked P2/P3a amplitude – reliably predicted the magnitude of bias. We synthesized these findings, as well as the other findings described so far, into a dynamical attractor model of visual WM.

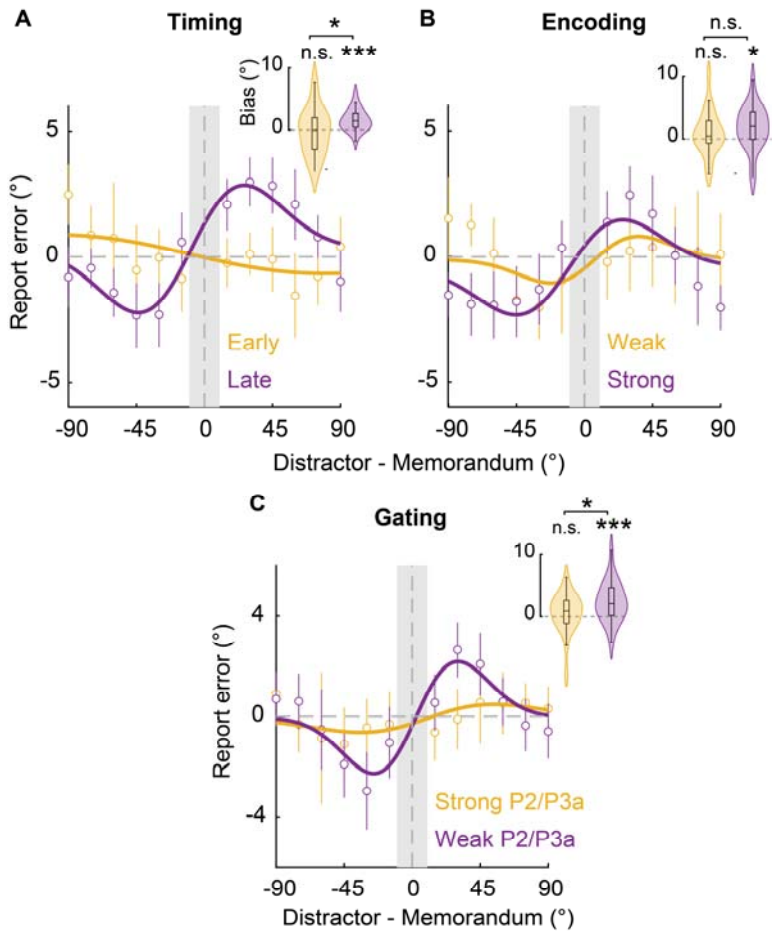



**Figure 4. Distractor's timing, encoding and gating predict mnemonic biases.**

**A.** Behavioral bias curves for the cued memorandum, but separately for trials in which the distractor appeared early (yellow) or late (purple) during the delay period. Data smoothed across sliding windows of three angle bins for clarity of visualization. (*Inset*) Violin plots showing the distribution of bias strengths across participants (n=23) for early (yellow) and late (purple) distractor trials. Other conventions are the same as in Figure 1F.

**B.** Same as in panel A, but for trials with weak (yellow) or strong (purple) distractor encoding. Other conventions are the same as in panel A.

**C.** Same as in panel A, but for trials with strong (yellow) or weak (purple) distractor-evoked P2/P3a amplitude, a putative marker of distractor gating (Fig. 3C). Other conventions are the same as in panel A. (All panels) *: $p<0.05$; **: $p<0.01$; ***: $p<0.001$; n.s.: not significant.

**An attractor model explains distractor-induced biases in visual WM**

To provide a parsimonious, mechanistic explanation of these distractor-induced biases, we designed a two-tier bump attractor model. Based on extensive recent literature[3–5,9,13,18,34–38], working memory maintenance in our model involves both the sensory (visual) cortex (VC), and higher cortex (HC), such as the prefrontal cortex (PFC) or the parietal cortex (PPC). In fact, the latter regions both are reported to be involved in robust maintenance in the presence of distractors[9,23,24,39]. Moreover, we include a key element of novelty in our model: while distinct attractors independently encode stimuli in the two hemifields, they interact with each other via competitive, cross-hemifield inhibition, based on biological observations[40–44]. We provide a brief description of the model below; a detailed description is provided in the Methods.

The VC and the HC were modeled with separate ring attractors; we also employed distinct attractors for encoding the memorandum in each (cued, uncued) hemifield, yielding a total of 4 attractor networks (Fig. 5A, top). Each attractor network comprised both excitatory (E) and inhibitory (I) neurons (512 each) with dynamics described by rate-based models (Methods, equations 1-2). Recurrent connections among E neurons within an attractor were local and topographic (Fig. 5A, bottom), whereas all other connection types (E-I, I-E, I-I) were global (see Methods). VC neurons provided input to HC neurons via topographic, excitatory feedforward connections and received input from HC via reciprocal, topographic feedback connections (Methods, equations 3-4). Based on our experimental observation, we expected cued information in the VC to be more robust to distractors (Fig. 2G). Therefore, HC→VC feedback connections were 20% stronger for the cued, than the uncued, hemifield (Fig. 5A, thick HC-VC connection). In addition, previous studies suggest that mnemonic information in higher order areas (e.g., PFC, PPC) is relatively robust to distractors[8,9,13,23,39]. Consequently, we transiently blocked feedforward connections from the VC to the HC during the maintenance epoch, thereby

rendering the HC mnemonic information immune to disruption by the distractor upon its presentation. Finally, based on previous reports of competitive visual inhibition across the hemifield[40–44], we also modeled cross-hemispheric inhibitory connections between the VC attractors in each hemisphere (Fig. 5A, black circles; equation 3); these topographic connections played a critical role in explaining the antagonistic pattern of biases (see next).

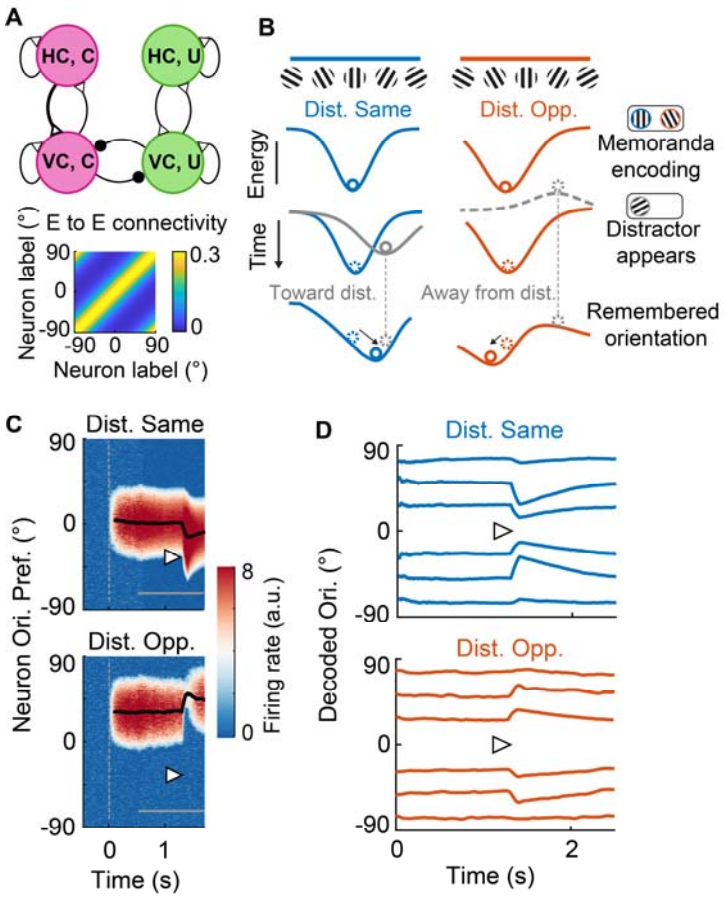



**Figure 5. Two-tier attractor model explains spatially antagonistic, mnemonic biases.**

**A.** (*Top*) Two-tier ring attractor model. The visual cortex (VC, bottom) and higher cortex (HC, top) were modeled with separate ring attractors. Distinct attractors also encoded stimuli in the cued (C, left, magenta) and the uncued (U, right, green) hemifields. Top-down input from HC to VC was stronger for the cued (thick black line) compared to the uncued side. Mutual inhibition between the VC,C and VC,U attractors was mediated by topographic, cross-hemispheric

inhibitory connections (black circles with horizontal arcs). (*Bottom*) Recurrent excitatory connectivity matrix, showing local and topographic connection weights (see text for details).

**B.** An energy landscape explains antagonistic distractor-induced biases. (*Top*) The memorandum (blue or orange circle) is maintained at the lowest energy state (highest firing rate). (*Middle, left*) When the distractor (gray circle) appears in the same hemifield, it causes a dip (energy minimum; solid gray line) in the energy landscape. (*Middle, right*) By contrast when the distractor appears in the opposite hemifield, it causes a peak (energy maximum; dashed gray line) in the energy landscape. (*Bottom*) As a result, the final energy landscape, obtained by summing the memorandum and distractor energies, biases the memorandum's representation *towards* the distractor's orientation when the latter appears in the same hemifield (blue circle) and biases the memorandum's representation *away* from the distractor's orientation when the latter appears in the opposite hemifield (orange circle).

**C**. Firing rate heat maps of the neural activity; hotter colors indicate higher activity. y-axis: E neurons arranged in order of orientation preference on the ring; x-axis: Time. (*Top)* In this example, the onset (dashed vertical line) of the memorandum gratings (0° orientation) causes a localized bump in firing rate at the respective location on the ring, which persists throughout the delay period (gray horizontal bar). Thick black curve: orientation decoded from the population activity vector. When the distractor grating (black triangle; -30° orientation) is presented transiently, and in the same hemifield as the memorandum, the latter's representation is biased toward the distractor's orientation. *(Bottom)* Same as in the top panel, but in this case the distractor grating (-30° orientation) is presented transiently, and in the opposite hemifield as the memorandum (+30° orientation). Here, the memorandum's representation is biased away from the distractor's orientation.

**D.** Decoded neural representations across time for six different orientations of the memoranda (curves, distinct simulations). Distractor grating (black triangle; 0° orientation) presented in the

same (*top,* blue curves) or the opposite (*bottom,* red curves) hemifield as the respective memoranda.

With this model, we sought to replicate the attractive and repulsive biases that occurred across space. We explain the expected sequence of events, first, with an energy landscape model (Fig. 5B); detailed simulations are presented subsequently. The energy landscape provides schematic representation activity and attractors in the network[45–47]: Minima (valleys) in the energy landscape reflect stable fixed points that cause stored patterns to drift towards their location whereas, conversely, maxima (peaks) reflect unstable fixed points that cause stored patterns to drift away from them.

The presentation of a memorandum produces a valley in the VC attractor's energy landscape at its orientation, which persists during the maintenance epoch due to recurrent excitation (Fig. 5B, top). Next, the distractor appearance in the same hemifield as the memorandum (DS) causes another dip in the VC attractor's energy landscape at the distractor's orientation (Fig. 5B, left column, middle row, gray trace). If the dip induced by the distractor is large enough, it perturbs the minimum in the landscape causing the minimum to shift, and the maintained representation to drift, towards the distractor's orientation (Fig. 5B, left column, bottom row, blue trace); this yields an attractive bias in the distractor's hemifield (Fig. 5B, left column, bottom row, blue circle). By contrast, the distractor also creates a peak at its own orientation in the VC attractor's energy landscape in the opposite hemifield (DO); this energy peak arises as a result of the topographic, cross-hemispheric inhibitory connections (Fig. 5B, right column, middle row, gray trace). As schematized, this peak repels the minimum in the landscape, and causes the maintained representation of the memorandum to drift away from the distractor's orientation (Fig. 5B, right column, bottom row, orange trace); this yields a repulsive bias in the hemifield opposite the distractor (Fig. 5B, right column, bottom row, orange circle).

Next, we describe the results of simulating the model (parameters in SI Table S1). We modeled the initial memory array presentation with an additive input current to the VC-E neurons of each attractor (cued, uncued) using a von Mises spatial profile centered on the orientation of the respective grating ($\kappa_M$=2.7°, equation 5). This initiated localized bumps of activity in VC attractors, at locations corresponding to the orientations of the encoded gratings (Fig. 5C top and bottom, activity at 0° and +30°, respectively). Upon disappearance of the memory array, during the delay period (Fig. 5C, gray underbar), the bump of activity was maintained through recurrent excitation in each attractor[26,27,48] (Fig. 5C, red patches). A distractor was presented at t=1300 ms during the delay period (Fig. 5C, triangle); distractor presentation was modeled, also with a von Mises input profile ($\kappa_D$=2.3°, equation 6) centered on the distractor's orientation (Fig. 5C; distractor at –30° for both panels).

Upon its presentation the distractor induced antagonistic mnemonic bias across hemifields (Fig. 5C). When the distractor was presented in the same hemifield as the memorandum the bump of activity shifted systematically toward the distractor's orientation (Fig. 5C, top, heat map). By contrast, when it was presented in the opposite hemifield as the memorandum the bump of activity shifted systematically away from the distractor's orientation (Fig. 5C, bottom, heat map). We repeated these simulations with six different memoranda of different orientations, with a distractor always appearing at 0° orientation. In each case, the distractor biased the decoded neural representations towards (Fig. 5D, top) or away from (Fig. 5D, bottom) itself, albeit to varying degrees (see next).

We quantified the neural bias for each of these distractor conditions (Methods); briefly, decoded orientations were estimated as the peak of the activity bump in the VC attractor. In line with the experimental findings, the distractor produced an attractive mnemonic neural bias when it appeared in the same hemifield (Fig. 6A) and a repulsive mnemonic bias when it appeared the opposite hemifield (Fig. 6B). The magnitude of this bias varied parametrically with the angular difference between the distractor and the memoranda: the strongest bias appearing for

intermediate angle differences, and the weakest appeared for extreme differences (0° or ±90°). Similarly, with behavioral reports, the distractor produced a strong, attractive bias in its hemifield (Fig. 6C); however, it produced a repulsive, albeit considerably weaker, bias in its opposite hemifield (Fig. 6D). This disparity in bias strengths occurred also in the experimental data across hemifields (Fig. 1 F, G). In our model this can be explained by the fact that cross-hemifield inhibition decayed comparatively faster in the opposite hemifield, aided by error-correcting dynamics arising from the HC attractor (see next).

In addition, these simulations replicated key experimental observations regarding the effects of distractor features on behavioral bias magnitude (Methods). Briefly, behavioral estimates of orientation were obtained by averaging the memorandum's orientation decoded from both VC and HC at the end of the trial; this averaging rendered behavioral estimates more robust to noise (see SI Results Section for a detailed analysis of error-correcting dynamics).

First, in simulations, as in the experimental data, late distractors produced stronger biases (Fig. 6E). We propose that this occurred because the top-down feedback from HC to the VC mitigated distractor interference in the VC through error-correcting dynamics[8,9,24] (SI Results Section on *Error correcting dynamics*). Specifically, late distractors appeared later during the WM maintenance epoch and were temporally more proximal to the response probe than early distractors; consequently, the former received ameliorative top-down feedback for a shorter duration than the latter. This produced a higher "residual" bias for late (Fig. 6E, purple) compared to early (Fig. 6E, yellow) distractor trials at the end of the delay epoch, matching our experimental findings.

Second, strongly encoded distractors produced stronger biases in the simulations also (Fig. 6F). Here, we modulated distractor encoding strength by affording a 60% higher input current amplitude to strongly, as compared to weakly, encoded distractors. The stronger perturbation of the energy landscape induced by the strongly encoded distractors yielded faster

drifts in the memorandum's representation thereby yielding a higher bias for strongly (Fig. 6F, purple) compared to weakly (Fig. 6F, yellow) encoded distractors; again, these results matched our experimental observations.

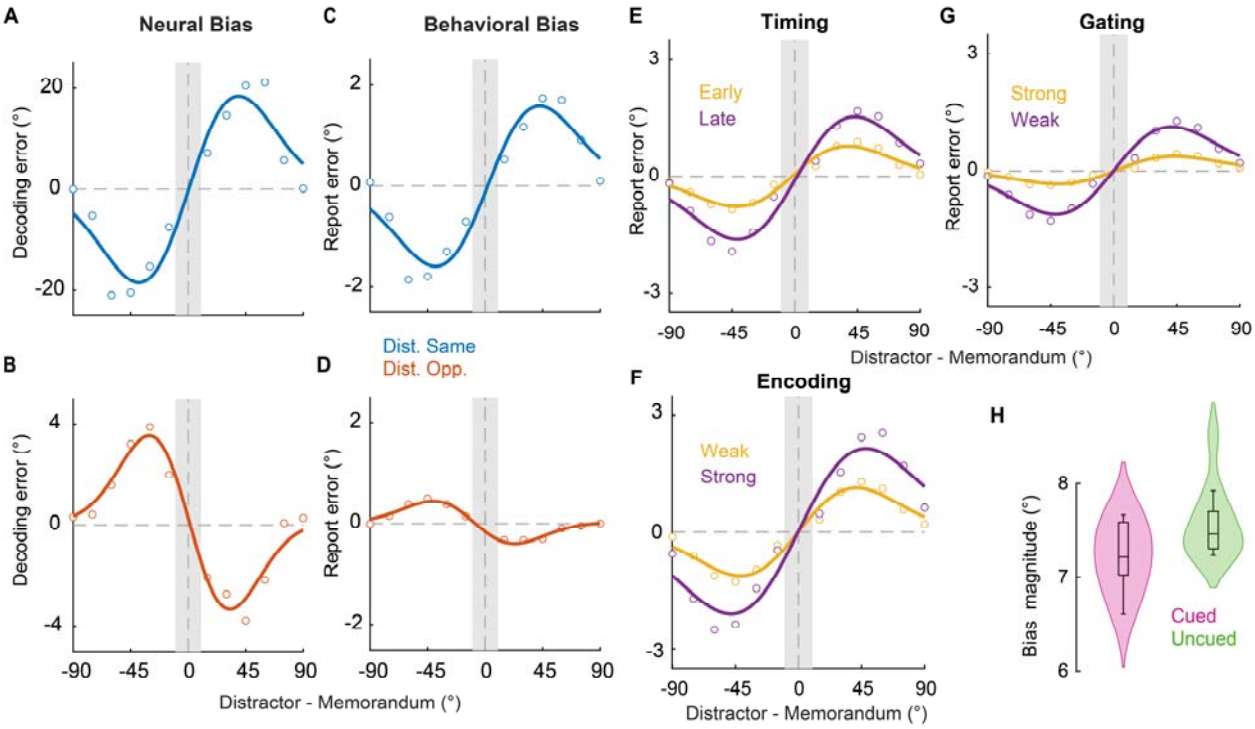



**Figure 6. Model simulations replicate distractor-induced biases in working memory.**

**A-B.** The model's neural bias measured 100 ms following the distractor onset when the distractor appeared in the same (A) or opposite (B) hemifield as the respective memorandum.

**C-D.** The model's behavioral bias measured 100 ms following the distractor onset when the distractor appeared in the same (E) or opposite (F) hemifield as the respective memorandum.

**A-D**. Data pooled across cued and uncued locations. Other conventions are the same as in Figures 1F-G.

**E-G.** Same as in Figures 4A-C, but showing the effect of distractor's timing (E), encoding strength (F) and input gating (G) on the model's behavioral bias for the cued memorandum. Other conventions are the same as in Figures 4A-C.

**H.** Magnitude (absolute value) of model's neural bias for the cued (magenta) versus the uncued (green) memoranda. Data averaged across distractor locations. Violin plots showing the distribution of values across 10 runs of the model (n=1000 simulations for each run); each simulation initialized with different random seeds, and values were averaged across simulations within a run.

Finally, weakly gated distractors produced stronger biases, again, in simulations (Fig. 6G). We modeled the extent of distractor gating by modulating feedback input from the HC to VC by 2x; we hypothesized that stronger feedback rendered mnemonic representations in the VC more robust to interference. Indeed, the magnitude of distractor-induced bias was systematically higher for the weakly-gated distractors (Fig. 6G, purple), compared to the strongly gated distractors (Fig. 6G, yellow), consistent with experimental results (see also SI Fig. S4). Interestingly, the effect of this gating mechanism was also apparent in our model, in another way. Due to stronger top-down feedback from HC to VC attractor on the cued hemifield, error-correcting dynamics were stronger in the cued hemifield. This led to stronger distractor gating and a weaker neural bias for the cued (Fig. 6H, magenta) compared to the uncued (Fig. 6H, green) memorandum in the model's neural representation.

Taken together, our results uncover space-specific, antagonistic patterns of distractor-induced biases in visual working memory. Behavioral reports, and the mnemonic neural representation of the memorandum, were biased towards the distractor's orientation when the latter appeared in the same hemifield as the former. By contrast, the mnemonic neural representation was biased away from the distractor's orientation when the latter appeared in the opposite hemifield. Cued neural representations were more strongly shielded from distractor interference as compared to uncued representations. Our two-tier ring attractor model parsimoniously replicated each of these experimental observations. Moreover, the model

comprehensively explained diverse factors that mediate distractor-induced bias, including distractor timing, strength of encoding and input gating. Overall, the results identify putative neural underpinnings of space-specific distractor biases in visual working memory.

## Discussion

We discovered that a salient sensory distractor that shares task-relevant features with the memoranda produces space-specific, antagonistic biases in visual working memory. The memorandum's neural representation was biased towards the distractor's orientation when the latter occurred in the same hemifield as the former, but was biased away when these stimuli appeared in opposite hemifields. [22,49–52]Over time, mnemonic representations became more labile and prone to interference by distractors that appeared later in the maintenance epoch. Yet, information associated with the selected (retro-cued) hemifield was held more robustly and shielded against distractor interference, possibly by "error-correcting" top-down control mechanisms[8,9,13,19,24].

Several lines of evidence confirm that the distractor biased the information maintained in WM (mnemonic bias), rather than merely inducing a bias during the behavioral report (response bias). First, had the distractor merely produced a response bias, we would have expected such a bias to occur regardless of the distractor's location. In our task, even though the response probe and bar were always presented centrally, we observed distinct patterns of biases – either toward the distractor's orientation, or a null bias (Fig. 1F-G) – depending on the distractor's location relative to the probed memorandum. Second, the distractor was flashed briefly (100 ms) and was followed immediately by a noise mask (pedestal) to diminish its persistent representation. Moreover, participants were aware that the distractor was task irrelevant, and it appeared at least 500 ms before the probe, rendering biases arising from persistence in iconic memory unlikely. Third, and most importantly, decoding the neural representation of the memorandum during the maintenance epoch revealed distinct, space-specific, antagonistic patterns of distractor-induced biases, providing direct evidence for distractor-induced biases in the representations maintained in WM.

The pattern of biases observed in the neural representations of the memoranda cannot be readily explained by distractor-evoked activity “leaking” into the decoding analysis. For the bias analysis, neural data were analyzed in a time window centered at 350 ms after distractor onset, to avoid distractor-evoked potentials from strongly affecting memorandum decoding; shifting this window by an additional +100 ms yielded similar results. In addition, a cross-decoding analyses (SI Fig. S3C) revealed no significant evidence for cross-generalized decoding between the distractor and memorandum’s neural representations. In addition to these, several lines of evidence in our study render such a confound unlikely. First, although information leakage from the distractor can produce an attractive bias, it cannot readily account for the repulsive bias – away from the distractor’s orientation – for the memorandum in the opposite hemifield. Second, the pattern of biases was strong for the uncued, but not the cued memorandum (SI Fig. S2); again, a direct leakage of distractor information into the decoding analysis cannot differentially affect the decoding of memoranda in the different hemifields. Finally, we observed distractor-induced biases in behavior also, indicating that biases we observed in the memorandum’s decoded representation may form the neural basis for distractor interference in working memory.

Bias trends in behavioral reports matched those in the neural representations, barring two key differences. First, while both behavioral and neural data showed attractive biases, evidence for repulsive biases was lacking in behavioral reports (Fig. 1F), or weak at best (SI Fig. S1B), when the distractor appeared in the opposite hemifield as the memorandum. We propose that the repulsive bias induced by the distractor in the hemifield opposite the memorandum decayed more rapidly than the attractive bias induced in the same hemifield, as a result of which it did not produce a measurable impact on behavior. Our modeling suggests that the faster decay of the repulsive bias reflects more labile effects of cross-hemifield inhibition. Second, behavioral reports in the cued hemifield were biased toward the distractor, even though

the cued neural representation did not exhibit such a bias (SI Fig. S2B). It is possible that biases in the cued representation occurred in activity-silent states, mentioned previously, or in regions outside of occipital cortex, which were not accessible to our decoding analysis. Future work, involving neural recordings in other regions (e.g., parietal cortex) may be necessary to identify the neural origins of the bias for the cued memorandum.

Our results provide concurrent validity to the sensory recruitment hypothesis of visual WM, which posits that brain regions involved in processing the memorandum during encoding also contribute to its maintenance in WM[1–5]. We were able to successfully decode both the orientation of the memoranda during the delay epoch, as well as distractor-induced biases in their neural representations from the same set of posterior electrodes (Fig. 2A-B). Yet, due to the low spatial resolution of EEG we cannot readily disambiguate whether the neural signatures that we observed originated entirely from the early visual cortex or also included contributions from other areas, like the parietal or prefrontal cortex[55]. Lastly, distractor-induced bias effects were hemifield-specific; this is consistent with previous studies which suggest that storage of mnemonic information involves a lateralized buffer including, possibly, also the sensory cortex [22,49–52].

Critically, prioritization in working memory promoted distractor resilience: the selected neural representation, associated with the cued memorandum, was significantly less biased by the distractor as compared to the uncued memorandum's representation. Taken together with previous studies[24,53], our results show that prioritized information is more robust to interference, possibly due to feedback-mediated stabilization of the cued memorandum's representation[24,53]. Indeed, when simulated in our computational model, top-down feedback from higher cortex stabilized WM representation in the visual cortex on the cued side (Fig. 6H). On the other hand, it is also possible that such stabilization of the cued representation arises from a mixture of persistent activity and activity-silent states[7,29,48,54]; the latter are arguably less prone to

interference from sensory distractors[7]. Future experiments can disambiguate these hypotheses by using a “pinging” protocol that activates latent representations and studying the effect of distractor interference on these activated representations.

Among the multiple electrophysiological indices we tested, we propose the P2/P3a ERP amplitude as a putative marker for distractor gating into working memory. Previous literature has often linked the P2 component to the gating (or suppression) of distractor information during working memory tasks[56–58]. For instance, higher distractor-evoked P2 amplitude, indicative of greater distractor gating, correlated positively with performance in an auditory WM task[31]. Similarly, the short peak latency P3 component, referred to as the P3a, possibly originates in the frontal lobe[57] and indexes top-down inhibition of distractors[56,57]. In fact, distractor-evoked P3a amplitudes were correlated with frontal lobe grey matter volume across individuals[59], suggesting its involvement in the distractor-related processing. Our study extends these findings and shows that the P2/P3a amplitude contralateral to the distractor was higher at the cued compared to the uncued hemifield; in other words, the P2/P3a may index a stronger input gating of cued mnemonic information stored in a lateralized mnemonic buffer (Fig. 3C). Moreover, the P2/P3a amplitude directly predicted the magnitude of the distractor-induced bias, providing further confirmation for this “gating hypothesis”.

We synthesized our experimental findings with a two-tier ring attractor WM model. Attractor models have been frequently used for explaining diverse phenomena in WM, including variable precision[26], serial biases[48], categorical biases[60] and the like[27]. For instance,–a recent study employed an attractor model to explain serial biases in spatial WM[48]. Other studies employing attractor models have been able to explain the mechanisms behind trial to trial variability and the presence of categorical biases in visual WM[26,27,48]. In our study, a ring attractor model was a natural candidate to explain our results. Recent research has shown that maintenance in visual working memory (WM) involves multiple regions across the brain,

including both sensory and higher-order cortex[18,35,37]. We, therefore, employed a two-tier attractor model that includes both sensory and higher order brain regions to simulate our experimental findings.

In our model, a higher-order cortical region (HC) stored information relevant to the memorandum and provided top-down influence to the visual cortex (VC) during the delay period to mitigate distractor interference in WM. Candidate regions for the HC include parietal cortex or prefrontal cortex both of which are known to contain robust WM representations, especially in the presence of distractors[9,13,23,24,39]. In our model, the representation in HC was rendered immune to distractor interference by transiently blocking its bottom-up input from VC during the delay period. Neurophysiological mechanisms that could mediate such dynamic disconnections include oscillatory coherence[61] or neuromodulatory gating (e.g., acetylcholine[62]). Yet, some recent studies have shown that disruption in visual WM maintenance occurs in both lower (V1, V4) and higher level cortex[11]. In this case, the coordinated activity across these regions could help improve the robustness of maintenance in WM, even in the presence of distractors[18,35–37,63]. Consistent with these results, our simulations showed that combining VC and HC activity yielded a more accurate behavioral readout of the memorandum's orientation, as compared to either VC or HC activity alone (SI Fig. S3B). Future work, employing fMRI, perhaps in combination with EEG, could help identify the precise brain regions, and their dynamics, that mediate distractor biases in WM.

Our model accurately reproduced the effect of distractor timing on mnemonic bias: distractors presented later during the delay epoch induced a stronger bias, and vice versa. This experimental result is in line with literature suggesting that neural WM representations are not static but evolve over time, as dynamic processes[8,30,64]. We leveraged these findings in our model to explain distractor timing effects. In our model, the bias due to error-correcting dynamics in visual WM decays steadily; these dynamics are, in turn, mediated by top-down

feedback from higher cortex (e.g., PFC[9,13,23,24,39]). Due to this effect, early distractor trials provided more time for the error-correcting mechanism to operate thereby yielding a weaker behavioral bias, as compared to late distractor trials. While model simulations accurately recapitulated experimental data, future experiments can confirm this hypothesis by causally manipulating the activity of higher order brain regions (through brain stimulation) during a visual WM task and studying its effects on the extent of error-correcting dynamics observed in lower order brain regions and in behavior**.**

In conclusion, our study offers putative mechanistic insights into neural processes by which distractors induce space-specific biases in working memory. Complementing these experimental results, our simulations suggest key computational hypotheses that by which distractors bias in WM representations. These hypotheses may be tested causally with non-invasive brain stimulation (e.g., transcranial magnetic stimulation) by focally perturbing activity in specific brain regions (e.g., visual cortex, prefrontal cortex), and at specific times during the delay epoch. Understanding the neural basis of distractor interference could have critical implications for improving the robustness of working memory in patients with attention deficit disorder (ADD) or mild cognitive impairment (MCI[65]).

# Materials and Methods

## Participants

24 participants (7 female; age range: 23-38 yrs; median age: 28 yrs), all right-handed, participated in the experiments. All participants had normal, or corrected-to-normal, vision and no known history of neurological disorders. Participants provided written, informed consent to take part in the experiments. Study protocols were approved by the Institute Human Ethics Committee at the Indian Institute of Science, Bangalore.

Sample size estimation. Sample sizes were not estimated with power analysis. The sample size in this study (n=24) is largely similar to that of several earlier studies on working memory and EEG-based grating orientation decoding[12,29,30].

## Behavioral data acquisition

Task and stimuli. Participants completed the experimental tasks in a dimly lit room while seated, with their heads secured by a chin rest, with the eyes located 60 cm from a stimulus display screen. Stimuli were displayed on a contrast calibrated (i1 Spectrophotometer, X-Rite Inc.) 24-inch LED monitor (XL2411Z, BenQ corp.) with a resolution of 1920x1080 pixels and a refresh rate of 100 Hz. Stimuli were programmed in Matlab 2015b (Mathworks Inc.) with Psychtoolbox version 3.0.15. During the experiment, participants used an optical mouse (M-U0026, Logitech) to provide responses. Participants' gaze position was tracked throughout the experiment using an infrared eyetracker (500 Hz, HiSpeed, iView X, SensoMotoric Instruments Inc.).

Each trial began with a fixation dot (full black, radius: 0.3 dva) presented at the center of the screen. Concurrently with this, circular placeholders (uniform white noise patches) were presented symmetrically in either hemifield along the azimuth (eccentricity: ±8.0 dva, radius: 4.0 dva); both the fixation dot and the placeholders remained on the screen throughout the trial. 500

ms after fixation dot onset, the memory array, comprising two sine-wave gratings (50% contrast, radius: 3.5 dva, spatial frequency: 0.9 cpd), appeared for a brief interval (100 ms), each embedded (50% alpha blended) concentrically inside one of the placeholders in each hemifield. 500 ms after memory array offset a retro-cue (white filled semicircle) appeared for 100 ms which indicated the location of the grating that would likely be probed for response (cue validity: 70%). Following a random delay (500-1250 ms, drawn from an exponential distribution), a singleton distractor grating (100% contrast) appeared for 100 ms in one of the two hemifields, pseudorandomly, with equal probability. The distractor's size and spatial frequency was identical to that of the memory array gratings, but its orientation was drawn randomly from a uniform distribution across all angles at least ±10° away from either grating's orientation in the memory array. The distractor was irrelevant to the task and the participants were explicitly instructed about the same. No distractor grating appeared in 20% of the trials, interleaved among the distractor presentation trials.

A fixed interval (1950 ms) after retro cue onset, a response probe (white or black filled circle, 0.3 dva, with a yellow ring around it) appeared on the screen indicating the grating relevant for response. A white (black) probe indicated that the participants must reproduce the orientation of the cued (uncued) grating. A randomly oriented bar (length: 7 dva, width 0.3 dva) appeared at the center of the screen upon response initiation. Participants reproduced the orientation of the grating on the probed side from memory by rotating the bar with the mouse using their right (dominant) hand, and pressed the left mouse button to register their response. Although there was no constraint on the time to initiate the response after probe onset, participants had to complete their response, once initiated, within 2 seconds. The next trial began after an inter-trial interval of 2 seconds.

<u>Training and testing.</u> Before the main experiment, participants completed between 2-4 blocks (100 trials each) of the task, typically, without EEG recordings. To familiarize and train

participants on the task, feedback was provided at the end of each trial. The feedback consisted of an oriented, magenta bar whose orientation precisely matched that of the probed grating shown along with the reported orientation (black bar); the feedback was shown for 200 ms at the end of the response period. Following this, they also received textual feedback, displayed on the screen, which indicated whether their response was “precise” (error <= 15°) or “imprecise” (error > 15°). Training data were not used for further analyses.

Following a brief break (typically 20 min), participants performed the main task, without behavioral feedback, but with concurrent EEG recordings (see next). Each participant completed 6 blocks of the task (100 trials each), with periodic breaks at the end of every set of 50 trials. Thus, in total, we acquired and analyzed 14,400 trials of behavioral data from 24 participants.

<u>Eye-tracking and gaze-based trial exclusion.</u> For all participants, gaze fixation position was monitored binocularly at 500 Hz, and stored offline for analysis. In *post hoc* analyses, any trial in which the subject’s gaze deviated by more than 1 dva from the fixation dot along the azimuthal direction was flagged, and excluded from further analysis. Given that we were primarily interested in the effect of distractors on visual working memory, this criterion for gaze-based exclusion was applied only in a window around (100 ms before to 400 ms after) distractor onset during the delay epoch. We did not reject trials based on gaze deviation during other times in the delay epoch, because, barring the fixation dot and placeholders, there were no other stimuli on screen during this epoch. We also did not reject trials based on gaze deviation during the memory array presentation, because there was no prior information (or pre-cue) that could have systematically biased gaze toward either grating. In addition, no gaze-based rejection was performed for trials in which the distractor did not appear. Finally, for 6/24 participants gaze fixation could not be reliably recorded, largely because these participants wore eyeglasses with

strongly reflective lenses; trial rejection rates for these participants were >15%. To avoid major biases in the statistical distribution of task parameters, for these 6 participants, we included all trials in the analyses; the results were similar even with these participants excluded from the analyses. For the remaining participants (n=18/24), the median eye-tracking rejection rate was 5.2±0.8% (mean±s.e.m.).

**Behavioral data analyses**

Estimating precision and bias in behavior. Participants' reported their best estimates of grating orientation. Simple behavioral metrics, computed directly from the data, revealed the effects of the distractor on precision and bias in participants' reports[66–68], without the need for fitting complex psychophysical models[69].

We computed the effect of the distractor on the precision of each participant's orientation report using the circular standard deviation (CSD)[68], a circular measure of dispersion of the orientation reports about its true value. A smaller CSD indicates higher average precision in orientation reports. Data were pooled from trials across all blocks for this analysis. We computed and compared the CSD separately for the trials in which the distractor appeared on the same side, versus the opposite side, of the probed stimulus; this analysis was performed separately for the cued and uncued grating reports.

We also computed the effect of the distractor in biasing orientation reports for the probed grating, following a standard protocol[28,67]. First, we plotted the median orientation report error for the probed grating (reported – actual probed orientation) as a function of the angle difference between the distractor and the probed grating (distractor – actual probed orientation); this computation was performed with trials pooled in overlapping bins (size: ±15°, shift: 1°; range: –90° to 90°), based on the difference between the distractor and the probed grating orientation.

Note that, in Figure 1E, if orientation reports were systematically biased towards the distractor orientation, then the y-axis (orientation report error) would be positive for positive values of the x-axis (relative distractor orientation), and vice versa for bias away from the distractor orientation. We quantified this bias by computing the signed area under the curve, with a procedure similar to that employed in previous studies (e.g., Myers et al[28]). Briefly, the area under the curve in quadrants II and IV were summed and subtracted from the summed area under the curve in quadrants I and III (Fig. 1E) – relative orientation values from -10° to 10° were excluded in this analysis because the distractor's orientation was at least 10° away from either of the memory array grating orientations (see Methods, Behavioral Data Acquisition section). This analysis was performed both for with data from all trials, and also separately for trials in which cued and uncued gratings were respectively probed. For visualization purposes, the bias curves were fit with a Gaussian derivative function[67].

**EEG data acquisition and preprocessing**

EEG acquisition. High density EEG data was acquired with the Active Two recording system (Biosemi, Amsterdam, Netherlands), with 128 Ag-AgCl active electrodes (Biosemi Inc.). During recording, the data were referenced online to the Common Mode Sense (CMS) active electrode. The data was recorded continuously throughout the task with a sampling rate of 4096 Hz and stored for offline analysis. EEG data for one participant could not be retrieved due to a critical hardware error. EEG analyses were performed with the remaining 23/24 participants.

EEG Preprocessing. EEG data was preprocessed in MATLAB using the EEGLAB[70], ERPLAB[71] and Noise Tools (http://audition.ens.fr/adc/NoiseTools/) toolboxes. The data was first resampled at 256 Hz (*pop_resample* function, EEGLAB), bandpass filtered from 0.05 to 40 Hz using

Butterworth filter of order-2 (*pop_basicfilter*, ERPLAB), and cleaned of line noise at 50 Hz (*pop_cleanline*, EEGLAB). Next, noisy electrodes were identified as those having more than 33% of their values >4x the median absolute value across all electrodes and time points (*nt_find_bad_channels,* Noise Tools); this yielded a channel removal rate of 4.1% ± 0.8% [mean ± s.e.m.] on average, per participant. The removed electrodes were then interpolated using the neighboring electrodes' data with spherical spline interpolation (*pop_interp,* EEGLAB). Subsequently, the trials were epoched from -1600 to 2300 ms relative to the cue onset and median re-referenced by subtracting the median value across all electrodes (Rolston et al., 2009). The epoched data were then baseline normalized by subtracting the mean activity from -200 to 0 ms relative to cue onset from all time points, separately for all electrodes and trials (*pop_rmbase,* EEGLAB). Next, we applied independent component analysis (ICA) using the infomax algorithm (*pop_runica*, EEGLAB) to identify artifacts arising from physiological and other, non-neural sources. We used ICLabel[72], an EEGLAB extension that probabilistically identifies ICs corresponding to neural sources versus those arising from eye movements, muscle artifacts, channel noise and other noise sources. Any component belonging to a non-neural source with high probability (>70%) was flagged, and removed after visual inspection; 3.2 ± 0.6 [mean ± s.e.m.] components were removed on average, per participant. In a final pass, we identified noisy trials using the SCADS algorithm[73]. Briefly, SCADS identifies outlier signal values based on statistical measures like the standard deviation, maximum amplitude, and the like. Trials in which SCADS flagged >20 percent of electrodes (outlier criterion, $\lambda$=5.0) were removed from the analysis. In other trials, noisy electrodes were linearly interpolated based on the nearest 4 neighbors; this yielded a median trial removal rate of ~3% trials per participant.

**EEG data analyses**

Decoding stimulus orientation. We decoded the orientation of memory array and distractor gratings based on multivariate pattern analysis of EEG responses. Whereas the memory array grating orientations were decoded both during memory array presentation (Fig. 1A) and the delay epoch (Fig. 1A), distractor grating's orientation was decoded only during its presentation epoch (Fig. 1A). The orientations of the memoranda were decoded with 14 posterior electrodes over the hemisphere contralateral to the respective stimulus (Fig. 2B, inset) pooled across the cueing conditions (see also control analysis, next, in which occipital electrodes were used bilaterally for decoding). The distractor's orientation was decoded with all 39 posterior electrodes over occipital cortex (e.g., Fig. 3A-B). We employed a multivariate decoder based on the Mahalanobis-distance, following a procedure established in literature[12,29,30]. Briefly, EEG data were downsampled to 128 Hz by taking an average of each set of two successive samples (original sampling rate: 256 Hz). To improve the decoding accuracy, we used a sliding window (size: 300 ms, shift: 8 ms) and concatenated all the data points in that time window from all electrodes into a one-dimensional vector (38 time points in each time window x number of electrodes). We utilized the pattern of temporal variability of the EEG response within each time window for decoding[30]. Therefore, we subtracted the mean activity across the time window for each channel, and applied principal component analysis (PCA) to reduce the number of feature dimensions; the top-k PCA dimensions that explained 95% of the variance across trials were retained for decoding ("encoding vector").

For training and evaluation, we employed a leave-one-trial-out cross-validation approach. For each participant, the model was trained on all but one trial, with orientations binned relative to the orientation in the left-out trial into one of 16 different bins (bin centers: 11.25° apart, bin width: ±30°). Then the Mahalanobis distance was computed between the test trial's encoding vector and the mean encoding vector of each training bin. The computed distances were de-meaned and the sign of the distances was reversed for ease of visualization.

To obtain a smooth decoding estimate, this procedure was repeated 8 times by sliding the bin centers by 1.4° each time, and the resultant decoding profiles were averaged across repetitions (e.g., Fig. 2A). To estimate decoding accuracy, the sign-reversed distances were multiplied elementwise by a cosine kernel, symmetric around 0° (period: 180°), and averaged. This procedure was performed for each participant, memory array location (left/right), distractor location (left/right), and each time window, separately.

<u>Estimating distractor-induced bias in neural representations.</u> To compute the distractor-induced bias in the neural representation of the memory array gratings, we first decoded the orientation of the memoranda. For this, first, we smoothed the decoded tuning curves using a moving average window (size: 80 ms) from 450 ms before to 750 ms after distractor onset on each trial. The mean orientation of the bin corresponding to the peak of the tuning curve was taken to be the decoded orientation for that time point and trial. Following this, the neural bias was computed from the decoded orientation in a way similar to the behavioral bias (see Methods section on *Estimating precision and bias in behavior*). Briefly, we plotted the median error in memorandum's orientation decoding (decoded – actual orientation) as a function of the angle difference between the distractor and the grating's actual orientation (distractor – actual orientation). Similar, to the behavioral bias, this computation was performed with trials pooled in overlapping bins (size: ±15°, shift: 1°; range: -90° to -90°), based on the difference between the distractor and the grating's actual orientation. We repeated this procedure separately for the distractor-same (decoded grating and the distractor appeared in same hemifield) and distractor-opposite (decoded grating and the distractor appeared in opposite hemifields) conditions, with all trials, and also separately for the cued and uncued gratings. The value of the decoding error computed in a ±150 ms time window centered at 350 ms following distractor onset was used to estimate the neural bias (Fig. 2C). This window was chosen to avoid the distractor-evoked

sensory response from confounding orientation estimates of the memoranda; shifting this window by up to ±100 ms yielded largely similar results.

Control analyses for neural bias estimation. We performed three control analyses to confirm the robustness of the neural bias effects. First, we tested whether the memorandum could be decoded during the delay period in the vicinity of distractor onset, to enable estimating neural bias induced by the distractor. But because the distractor biased mnemonic neural representations strongly (Fig. 2C), we did not expect to get superlative decoding accuracies for the memoranda around the time of distractor onset. As an alternative, therefore, we quantified the memorandum's decoding accuracy, pooled across the cued and uncued memoranda, in two windows: the first window spanned 500 ms window preceding the earliest distractor onset whereas the second window spanned 500 ms window post the final distractor offset. The decoding accuracies in these epochs over time are shown in SI Figure S3A-B. Second, we tested if distractor-related information could have leaked into the memorandum's decoded orientation. For this, we trained separate Mahalanobis-distance-based decoders based on each memorandum's orientation (left and right gratings) on each trial, and predicted the distractor's orientation on that trial, averaged across the two decoders; a higher than chance decoding accuracy would indicate that decoding cross-generalizes between the memory items and the distractor. The results of this cross-generalized decoding analysis are shown in SI Figure S3C. For these analyses, we employed on trials in which the distractor orientation was at least 45 degrees away from the memorandum's orientation were used, to avoid confusion between the actual memorandum and distractor's decoded representations on trials in which these stimuli were actually of similar orientations. Third, while in the main neural bias analyses (Fig. 2) we had employed only the electrodes in the hemisphere contralateral to the memorandum to decode its orientation, we now repeated the same analysis except that we employed 17

electrodes over bilateral occipital parietal cortex, whose positions were closest to the electrode configuration described in recent work[29] (SI Fig. S2D).

Distractor-evoked ERP estimation. The event related potential (ERP) was measured by signal-averaging across the same set of 14 contralateral (or ipsilateral) electrodes as employed in the decoding analyses. First, the data were high-pass filtered with a lower cut-off at 0.5 Hz using an FIR filter (N=500). Subsequently, the data were epoched from 800 ms before to 700 ms after distractor onset and then averaged across channels and trials to obtain the ERP waveform. This was done separately for each of the four distractor conditions (cued/uncued x left/right hemifield). Consistent with previous studies[31,33], the N1 amplitude was defined as the mean signal value in a window extending from 150 to 200 ms following distractor onset and P2/P3a amplitude as the mean signal value in a ±40 ms window around the respective P2 and P3a peaks.

Median-split analyses based on distractor encoding strength, timing and gating. We computed a measure of association between the strength of neural distractor encoding and behavioral bias. For this analysis, the strength of distractor encoding was quantified as the value of decoded cosine similarity obtained in a window from 100-400 ms after distractor onset. Then, we divided trials into “weak” and “strong” distractor encoding trials based on a median split of the cosine similarity metric; the median split was performed for each participant and for each distractor condition (cued/uncued x left/right hemifield) separately. Weak and strong distractor encoding trials were then aggregated across the conditions, and the behavioral bias in the cued orientation report was computed separately for the weak and strong trials with a procedure identical to that described previously (see Methods section on *Estimating precision and bias in*

*behavior*). The association between other metrics like the distractor timing or the strength of gating and behavioral bias was computed in an identical manner, except that the median split was performed based on the respective metric – time of distractor onset relative to the cue (early versus late) or amplitude of the distractor-evoked P2/P3a component over the electrodes contralateral to the distractor (weak versus strong gating). Statistical tests were performed both to assess whether the bias for each half of the median split was, individually, significantly different from zero, and also to assess for within-participant (pairwise) differences across the median split halves. For each split, we computed one-tailed Bayes Factors (see next), which enabled us to quantitatively evaluate evidence for or against the effect of each factor on distractor-induced bias.

**Computational model of distractor biases in WM**

<u>i) Network architecture.</u> We synthesized our behavioral and neural findings by modifying and extending a bump attractor model of working memory[12]. The visual cortex (VC) and higher cortex (HC), for each of the cued and uncued visual hemifields, were modeled with distinct ring attractors (total of 4 ring attractor models). Each attractor model comprised 1024 neurons in total out of which 512 were excitatory (E) and 512 were inhibitory (I). Each neuron was described by a standard rate model and represented stimulus orientation topographically, depending on its location along the ring (Fig. 5A; $\theta$ ranging from -90° to +90°).

For ease of illustration, we start by describing the dynamical equation governing the firing rate $r$ of the cued visual cortex attractor (VC,C) as a function of time $t$[26,27].

$$\tau_E \frac{dr_E^{VC,C}}{dt} = -r_E^{VC,C} + \phi\left(W_{EE}^{VC,C} r_E^{VC,C} - W_{IE}^{VC,C} r_I^{VC,C} + I_E^{VC,C} + C_E^{VC,C}\right) + \sigma_E^{VC,C}\xi(t) \qquad (1)$$

$$\tau_I \frac{dr_I^{VC,C}}{dt} = -r_I^{VC,C} + \phi\left(W_{EI}^{VC,C} r_E^{VC,C} - W_{II}^{VC,C} r_I^{VC,C} + C_I^{VC,C}\right) + \sigma_I^{VC,C}\xi(t) \qquad (2)$$

Where, $r_E^{VC,C}$ and $r_I^{VC,C}$ denote the firing rate of the excitatory and inhibitory neurons of the visual cortex, cued attractor (VC,C) respectively. $\mathrm{T}_E$ and $\tau_I$ are neuronal (integration) time constants for E and I neurons respectively, $W_{EE}^{VC,C}, W_{IE}^{VC,C}, W_{EI}^{VC,C}, W_{II}^{VC,C}$ are the connection strengths from E to E, I to E, E to I and I to I neurons respectively for the VC,C attractor. $C_E^{VC,C}$ and $C_I^{VC,C}$ are the constant input bias currents to E and I neurons respectively of VC,C attractor. $\Xi(t)$ represents Gaussian white noise with unit variance, scaled by parameters $\sigma_E^{VC,C}$ and $\sigma_I^{VC,C}$ for E and I neurons of VC,C attractor respectively. $I_{\mathrm{E}}^{VC,C}$ represents the input from other sources (explained in equations 3 and 4, see next). $\phi$ represents the function that maps the input current to the output firing rate adopted from Wimmer et al., 2014[26]

$$\phi(r) = \begin{cases} 0, & \forall\, r < 0 \\ r^2, & \forall\, 0 < r < 1 \\ sqrt(4r-3), & \forall\, r \geq 1 \end{cases}$$

Three of the four sets of recurrent connections – $W_{IE}, W_{EI}, W_{II}$ – were uniform and all-to-all (i.e., $w^{ij} = w_0\ \forall\, i, j$). On the other hand, recurrent excitatory connections $w_{EE}^{ij}$ were local, and followed a von Mises profile: $exp(\kappa cos(\theta_i - \theta_j))$ where $\kappa$ is the concentration parameter and $\theta_i$ and $\theta_j$ are the orientation preference of excitatory neurons $i$ and $j$ respectively.

The firing rates of three other attractors, namely, the visual cortex uncued attractor (VC,U), higher cortex cued attractor (HC,C) and higher cortex uncued attractor (HC,U) also follow identical dynamical equations. The values of the parameters for the 4 different attractor models are shown in SI Table S1.

Inputs to the cued visual cortex attractor ($I_{\mathrm{E}}^{VC,C}$, equation 1) comprised 3 contributions:

$$I_{\mathrm{E}}^{VC,C} = U^{VC,C} + W_E^{HC,C\rightarrow VC,C}\, r_E^{HC,C} + W_I^{VC,U\rightarrow VC,C}\, r_E^{VC,U} \qquad (3)$$

The first term, $U^{VC,C}$ represents the bottom-up (feedforward) inputs due to the external stimulus that become active during the presence of the memory array and distractor gratings (successively). The second term, $W_E^{HC,C\rightarrow VC,C} r_E^{HC,C}$ represents top-down (feedback) inputs from the HC to the VC attractor, organized topographically (one-to-one); here, $W_E^{HC,C\rightarrow VC,C}$ denotes the top-down excitatory connection weights from the HC,C to VC,C attractor. The third term, $W_I^{VC,U\rightarrow VC,C}\, r_E^{VC,U}$ represents long-range, lateral inhibitory inputs received from the VC attractor in the opposite hemisphere (cross-hemispheric inhibition); here, $W_I^{VC,U\rightarrow VC,C}$ represents the lateral inhibitory connection weights from the VC,U to VC,C attractor. Note that the magnitude this inhibition modulates with the firing rate of the excitatory neurons ($r_E$) in the opposite hemifield; an intervening inhibitory population that converts this long-range excitatory projection into functional inhibition (via weights, $W_I$) was not explicitly modeled. The inputs $I_E^{VC,U}$ to the uncued visual cortex attractor also followed an identical equation except for minor modifications to the respective bottom-up inputs ($U^{VC,U}$), top-down inputs ($W_E^{HC,U\rightarrow VC,U} r_E^{HC,U}$) and long-range, lateral inhibitory inputs ($W_I^{VC,C\rightarrow VC,U}\, r_E^{VC,C}$).

Inputs to the cued higher cortex attractor ($I_{\mathrm{E}}^{VC,C}$, equation 1) comprised only one contribution corresponding to bottom-up connections from the VC,C attractor, also organized topographically (one-to-one):

$$I_E^{HC,C} = W_E^{VC,C\rightarrow HC,C} r_E^{VC,C} \qquad (4)$$

Where $W_E^{VC,C\rightarrow HC,C}$ represents the topographic bottom-up connection strength from the E neurons of the VC,C to the E neurons of the HC,C attractor ($W_E^{VC,C\rightarrow HC,C}$). The inputs $I_{\mathrm{E}}^{HC,U}$ to the uncued higher cortex attractor also followed an identical equation except for its bottom-up inputs ($W_E^{VC,U\rightarrow HC,U}$).

<u>ii) Modeling the distinct task epochs.</u>

*Encoding:* 500 ms after the trial onset the memorandum was presented for a duration of 500 ms. The visual cortex (VC) attractor receives sensory input during this period (through $U^{VC,C}$ and $U^{VC,U}$ terms in equation 3). We modeled this by changing the spatial profile of the input current to the excitatory neurons of the VC attractor according to:

$$U_i^{VC,C} = A_M * exp(\kappa_M cos(\theta_i - \Theta_M^C)) \qquad (5)$$

Where, $U_i^{VC,C}$ is the input current to the excitatory neuron *I* of VC,C attractor, $A_M$ is the amplitude of the input current, $\kappa_M$ is the concentration parameter determining the spread, $\theta_i$ is the orientation preference of excitatory neuron $i$ and $\Theta_M^C$ is the memory array grating's orientation on the cued side. The spatial profile of the input current was centered on the E neuron representing the respective grating's orientation. This instantiated localized bumps of activity in VC, at locations corresponding to the orientations of the encoded gratings. A similar equation was used to model the input current for the VC,U attractor. The PFC attractor did not receive any direct inputs other than the bottom-up input received from the VC attractor during encoding (equation 4).

*Delay:* The delay period commenced immediately after memory array offset. During the delay period, until the end of the trial, we disconnected the bottom-up connections from the VC to the HC attractor ($W_E^{VC,C\rightarrow HC,C} = 0$, $W_E^{VC,U\rightarrow HC,U} = 0$ in equation 4), rendering the HC attractor immune to subsequent sensory input. 600 ms after the memory array onset, the cue appeared. To mimic prioritization of cued information in WM[24], we reduced the noise terms for both the cued attractors ($\sigma_E^{VC,C}, \sigma_I^{VC,C}, \sigma_E^{HC,C}$ $and$ $\sigma_I^{HC,C}$ in equations 1 and 2) by 60% and increased top-down feedback to cued VC attractor ($W_E^{HC,C\rightarrow VC,C}$ in equation 4) by 20%, following cue onset.

*Distractor presentation:* The distractor was presented for a duration of 0.1 seconds, at a random time (distributed exponentially) between 1200 ms and 1950 ms after the memory array onset, mimicking the behavioral paradigm. The distractor was presented either in the cued or the uncued hemifield with equal probability and was encoded by the visual cortex (VC) attractor representing the corresponding visual hemifield. Distractor presentation was modeled by adjusting the spatial profile of the input current to the excitatory neurons of the VC attractor as follows:

$$U_i^{VC,C} = A_D * exp(\kappa_D cos(\theta_i - \Theta_D)) \qquad (6)$$

Where $U_i^{VC,C}$ is the input current to the excitatory neuron $i$ of VC,C attractor, $A_D$ is the amplitude of the input current, $\kappa_D$ is the concentration parameter determining the spread, $\theta_i$ is the orientation preference of excitatory neuron $i$ and $\Theta_D$ is the distractor's orientation. In addition, we enhanced the cross-hemifield inhibitory weights ($W_I^{VC,U\rightarrow VC,C}$ or $W_I^{VC,C\rightarrow VC,U}$ in equation 3) multiplicatively (by ~30x) during distractor presentation; this enabled modeling the bias away from the distractor's orientation observed in the distractor-opposite condition (Fig. 2C, right).

iii) Quantifying simulated behavioral and neural metrics. We ran a total of 10,000 simulations with a different random seed each time; all results represent an average of these simulations. During each simulation, we randomly sampled the memory array (cued and uncued) and distractor orientation from a circular uniform random distribution. The "neural" orientation maintained by the network was decoded using the population vector representation in the visual cortex attractor given by:

$$\hat{\theta} = \arg\left(\sum_{j=1}^{N} r_{E,j}^{VC,C} e^{i\theta_j}\right) \qquad (7)$$

where, $\hat{\theta}$ is the decoded orientation from the VC,C attractor. $N$ is the number of E neurons. $r_{E,j}^{VC,C}$ and $\theta_j$ is the firing rate and preferred orientation respectively of neuron $j$ in the VC,C attractor. Orientations from uncued visual cortex attractor were also decoded in an identical manner, using its respective population vector. The model's neural orientation decoding was computed with the VC attractors alone to mimic the decoding of neural orientation from only the occipital electrodes in the EEG data. By contrast, drawing inspiration from recent work that has highlighted the involvement of both sensory (visual) cortex (VC) and higher brain areas (HC) in robust working memory maintenance[3–5,9,13,18,34], the decoded orientation from the VC and HC attractors at the end of the delay period were averaged to generate the model's behavioral orientation estimate (see SI Results Section); orientation from the higher cortex attractor were decoded in a manner identical to that in equation 7, using their respective population vectors. The model's neural bias was computed with the decoded orientation from the VC attractor's activity around the distractor onset, using a procedure identical to that described in EEG data analysis (see section on *Estimating distractor-induced bias in neural representations*).

**Statistical Analyses**

We employed a non-parametric Wilcoxon signed rank test for comparing median errors in orientation reports (Fig 1B, C). Permutation tests were performed for behavioral and neural bias across conditions; one-sided p-values are reported under the *apriori* hypothesis that the attractive (repulsive) biases occur when the distractor appeared on the same (opposite) side as the memorandum (Fig. 1F-G; 2C; 4A-C). All the other tests were two-sided, unless otherwise stated. To study the effect of distractor condition and cueing on precision in visual WM (Fig. 1C), we performed a 2-way ANOVA with the circular standard deviation as the dependent variable and distractor condition (distractor-same, distractor-opposite and no-distractor) and cueing

(cued and uncued) as predictors; participants was treated as random effects. *Post-hoc* analyses were performed by averaging the values across the cue conditions and comparing CSD values for distractor-same versus distractor-opposite conditions with Wilcoxon signed rank test. Similar analysis was performed for neural bias magnitude across cued and uncued conditions (Fig. 2D). To find significant time points in decoding accuracies plotted as a function of time (Fig. 2B), we used the two-sided cluster-based permutation test[75] with a cluster-forming threshold of $p<0.05$ and with 10,000 permutations of the data. The distribution of decoding accuracies across cued and uncued conditions was compared using permutation tests (two-sided) both for memorandum decoding and distractor encoding analysis, as well as for neural bias in distractor encoding (Fig. 3A,B). The distractor related ERPs across cued and uncued conditions were compared using the signed-rank test (Fig. 3C).

In addition, we also report the Bayes factor (BF[74]), a statistical measure that compares the likelihood of the alternate hypothesis to that of the null hypothesis under the observed data. BFs were computed with a JZS prior, using the Bayes Factor toolbox[76]. One-tailed BFs were computed specifically when comparing the effect of distractor features on behavioral or neural biases; other BFs were two-tailed, unless otherwise stated. Moreover, one-tailed BFs were also used to compare the strength of biases across the median splits, because we had clear, *apriori* hypotheses regarding the effect of each distractor characteristic (timing, encoding, gating) on bias magnitude. BFs computed based on a two-tailed versus one-tailed tests are reported as $BF_{10}$ versus $BF_{+0}/BF_{-0}$, respectively; the subscripts +0 and -0 denote the directionality of the hypothesis tested (greater or lesser than zero, respectively).

**Data and code availability**

Data and custom Matlab code to replicate the experimental findings as well as the computational modeling results are available at the following link:

https://osf.io/dqr7m/?view_only=71feeb114f744f8c987d6a22f230905f

**Declaration of generative AI and AI-assisted technologies in the writing process**

During the preparation of this work the second author used ChatGPT and Open AI as a language editing tool. After using this tool/service, the second author reviewed and edited the content as needed and takes full responsibility for the content of the parts of the publication that was co-written by them.

**Acknowledgements**

We thank Rishabh Bajpai and Mirudhula Mukundan for their help in piloting the task design. This research was supported by a Wellcome Trust-Department of Biotechnology India Alliance Intermediate fellowship, DST Swarna-Jayanti fellowship, a Pratiksha Trust Intramural grant, an India-Trento Programme for Advanced Research (ITPAR) grant and a Department of Biotechnology-Indian Institute of Science Partnership Program grant (all to DS).

**Declaration of Interest statement**

The authors declare no competing interests.

**Author contributions**

D.S. and S.G. designed the study; S.G. conducted the experiment; D.R and S.G. analyzed the data and performed model simulations; D.R, S.G. and D.S. wrote the paper.

## References


1. Pasternak, T. & Greenlee, M. W. Working memory in primate sensory systems. *Nat Rev Neurosci* **6**, 97–107 (2005).

2. Emrich, S. M., Riggall, A. C., La Rocque, J. J. & Postle, B. R. Distributed patterns of activity in sensory cortex reflect the precision of multiple items maintained in visual short-term memory. *J Neurosci* **33**, 6516–6523 (2013).

3. Gayet, S., Paffen, C. L. E. & Van der Stigchel, S. Visual Working Memory Storage Recruits Sensory Processing Areas. *Trends Cogn Sci* **22**, 189–190 (2018).

4. Scimeca, J. M., Kiyonaga, A. & D'Esposito, M. Reaffirming the Sensory Recruitment Account of Working Memory. *Trends Cogn Sci* **22**, 190–192 (2018).

5. Rademaker, R. L., Chunharas, C. & Serences, J. T. Coexisting representations of sensory and mnemonic information in human visual cortex. *Nature Neuroscience 2019 22:8* **22**, 1336–1344 (2019).

6. Bettencourt, K. C. & Xu, Y. Decoding the content of visual short-term memory under distraction in occipital and parietal areas. *Nat Neurosci* **19**, (2015).

7. Stokes, M. G. 'Activity-silent' working memory in prefrontal cortex: a dynamic coding framework. *Trends Cogn Sci* **19**, 394–405 (2015).

8. Lorenc, E. S., Mallett, R. & Lewis-Peacock, J. A. Distraction in Visual Working Memory: Resistance is Not Futile. *Trends Cogn Sci* **25**, 228–239 (2021).

9. Xu, Y. Reevaluating the Sensory Account of Visual Working Memory Storage. *Trends Cogn Sci* **21**, 794–815 (2017).

10. Lorenc, E. S., Sreenivasan, K. K., Nee, D. E., Vandenbroucke, A. R. E. & D'Esposito, M. Flexible coding of visual working memory representations during distraction. *Journal of Neuroscience* **38**, 5267–5276 (2018).

11. Hallenbeck, G. E., Sprague, T. C., Rahmati, M., Sreenivasan, K. K. & Curtis, C. E. Working memory representations in visual cortex mediate distraction effects. *Nature Communications 2021 12:1* **12**, 1–18 (2021).

12. Van Ede, F., Chekroud, S. R., Stokes, M. G. & Nobre, A. C. Decoding the influence of anticipatory states on visual perception in the presence of temporal distractors. *Nature Communications 2018 9:1* **9**, 1–12 (2018).

13. Xu, Y. Revisit once more the sensory storage account of visual working memory. *Vis cogn* **28**, 433–446 (2020).

14. Magnussen, S., Greenlee, M. W., Asplund, R. & Dyrnes, S. Stimulus-specific mechanisms of visual short-term memory. *Vision Res* **31**, 1213–1219 (1991).

15. Yoon, J. H., Curtis, C. E. & D'Esposito, M. Differential effects of distraction during working memory on delay-period activity in the prefrontal cortex and the visual association cortex. *Neuroimage* **29**, 1117–1126 (2006).

16. van Moorselaar, D., Gunseli, E., Theeuwes, J. & Olivers, C. N. L. The time course of protecting a visual memory representation from perceptual interference. *Front Hum Neurosci* **8**, (2015).

17. Rademaker, R. L., Bloem, I. M., De Weerd, P. & Sack, A. T. The impact of interference on short-term memory for visual orientation. *J Exp Psychol Hum Percept Perform* **41**, 1650–1665 (2015).

18. Christophel, T. B., Klink, P. C., Spitzer, B., Roelfsema, P. R. & Haynes, J. D. The Distributed Nature of Working Memory. *Trends Cogn Sci* **21**, 111–124 (2017).

19. Suzuki, M. & Gottlieb, J. Distinct neural mechanisms of distractor suppression in the frontal and parietal lobe. *Nature Neuroscience 2012 16:1* **16**, 98–104 (2012).

20. Berryhill, M. E., Chein, J. & Olson, I. R. At the intersection of attention and memory: The mechanistic role of the posterior parietal lobe in working memory. *Neuropsychologia* **49**, (2011).

21. Christophel, T. B., Iamshchinina, P., Yan, C., Allefeld, C. & Haynes, J. D. Cortical specialization for attended versus unattended working memory. *Nature Neuroscience 2018 21:4* **21**, 494–496 (2018).

22. Brincat, S. L. *et al.* Interhemispheric transfer of working memories. *Neuron* **109**, 1055-1066.e4 (2021).

23. Parthasarathy, A. *et al.* Mixed selectivity morphs population codes in prefrontal cortex. *Nature Neuroscience 2017 20:12* **20**, 1770–1779 (2017).

24. Myers, N. E., Stokes, M. G. & Nobre, A. C. Prioritizing Information during Working Memory: Beyond Sustained Internal Attention. *Trends Cogn Sci* **21**, 449–461 (2017).

25. Feldmann-Wüstefeld, T. & Vogel, E. K. Neural Evidence for the Contribution of Active Suppression During Working Memory Filtering. *Cerebral Cortex* **29**, (2019).

26. Wimmer, K., Nykamp, D. Q., Constantinidis, C. & Compte, A. Bump attractor dynamics in prefrontal cortex explains behavioral precision in spatial working memory. *Nature Neuroscience 2014 17:3* **17**, 431–439 (2014).

27. Compte, A., Brunel, N., Goldman-Rakic, P. S. & Wang, X. J. Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. *Cerebral Cortex* **10**, (2000).

28. Myers, N. E., Chekroud, S. R., Stokes, M. G. & Nobre, A. C. Benefits of Flexible Prioritization in Working Memory Can Arise Without Costs. *J Exp Psychol Hum Percept Perform* **44**, 398 (2018).

29. Wolff, M. J., Jochim, J., Akyürek, E. G. & Stokes, M. G. Dynamic hidden states underlying working-memory-guided behavior. *Nature Neuroscience 2017 20:6* **20**, 864–871 (2017).

30. Wolff, M. J., Jochim, J., Akyürek, E. G., Buschman, T. J. & Stokes, M. G. Drifting codes within a stable coding scheme for working memory. *PLoS Biol* **18**, e3000625 (2020).

31. Lijffijt, M. *et al.* P50, N100, and P200 sensory gating: Relationships with behavioral inhibition, attention, and working memory. *Psychophysiology* **46**, 1059 (2009).

32. Wongupparaj, P., Sumich, A., Wickens, M., Kumari, V. & Morris, R. G. Individual differences in working memory and general intelligence indexed by P200 and P300: A latent variable model. *Biol Psychol* **139**, 96–105 (2018).

33. Hillyard, S. A. & Anllo-Vento, L. Event-related brain potentials in the study of visual selective attention. *Proc Natl Acad Sci U S A* **95**, 781–787 (1998).

34. Pasternak, T. & Greenlee, M. W. Working memory in primate sensory systems. *Nat Rev Neurosci* **6**, 97–107 (2005).

35. Sreenivasan, K. K., Vytlacil, J. & D'Esposito, M. Distributed and dynamic storage of working memory stimulus information in extrastriate cortex. *J Cogn Neurosci* **26**, 1141–1153 (2014).

36. Mejías, J. F. & Wang, X. J. Mechanisms of distributed working memory in a large-scale network of macaque neocortex. *Elife* **11**, (2022).

37. Voitov, I. & Mrsic-Flogel, T. D. Cortical feedback loops bind distributed representations of working memory. *Nature 2022 608:7922* **608**, 381–389 (2022).

38. Khona, M. & Fiete, I. R. Attractor and integrator networks in the brain. *Nature Reviews Neuroscience* vol. 23 Preprint at https://doi.org/10.1038/s41583-022-00642-0 (2022).

39. Sakai, K., Rowe, J. B. & Passingham, R. E. Active maintenance in prefrontal area 46 creates distractor-resistant memory. *Nat. Neurosci.* **5**, 479–484 (2002).

40. Wyatte, D., Herd, S., Mingus, B. & O'Reilly, R. The Role of Competitive Inhibition and Top-Down Feedback in Binding during Object Recognition. *Front Psychol* **3**, (2012).

41. Dakos, A. S., Walker, E. M., Jiang, H., Stein, B. E. & Rowland, B. A. Interhemispheric visual competition after multisensory reversal of hemianopia. *European Journal of Neuroscience* **50**, 3702–3712 (2019).

42. Szczepanski, S. M. & Kastner, S. Shifting Attentional Priorities: Control of Spatial Attention through Hemispheric Competition. *Journal of Neuroscience* **33**, 5411–5421 (2013).

43. Reynolds, J. H., Chelazzi, L. & Desimone, R. Competitive Mechanisms Subserve Attention in Macaque Areas V2 and V4. *Journal of Neuroscience* **19**, 1736–1753 (1999).

44. Chinchani, A. M. *et al.* Tracking momentary fluctuations in human attention with a cognitive brain-machine interface. *Communications Biology 2022 5:1* **5**, 1–17 (2022).

45. Stokes, M. G. 'Activity-silent' working memory in prefrontal cortex: A dynamic coding framework. *Trends in Cognitive Sciences* vol. 19 Preprint at https://doi.org/10.1016/j.tics.2015.05.004 (2015).

46. Panichello, M. F., DePasquale, B., Pillow, J. W. & Buschman, T. J. Error-correcting dynamics in visual working memory. *Nat Commun* **10**, (2019).

47. Renart, A., Song, P. & Wang, X. J. Robust spatial working memory through homeostatic synaptic scaling in heterogeneous cortical networks. *Neuron* **38**, (2003).

48. Barbosa, J. *et al.* Interplay between persistent activity and activity-silent dynamics in the prefrontal cortex underlies serial biases in working memory. *Nat Neurosci* **23**, 1016–1024 (2020).

49. Delvenne, J. F. The capacity of visual short-term memory within and between hemifields. *Cognition* **96**, B79–B88 (2005).

50. Umemoto, A., Drew, T., Ester, E. F. & Awh, E. A bilateral advantage for storage in visual working memory. *Cognition* **117**, 69–79 (2010).

51. Adam, K. C. S., Robison, M. K. & Vogel, E. K. Contralateral Delay Activity Tracks Fluctuations in Working Memory Performance. *J Cogn Neurosci* **30**, 1229 (2018).

52. Luria, R., Balaban, H., Awh, E. & Vogel, E. K. The contralateral delay activity as a neural measure of visual working memory. *Neurosci Biobehav Rev* **62**, 100 (2016).

53. Souza, A. S. & Oberauer, K. In search of the focus of attention in working memory: 13 years of the retro-cue effect. *Atten Percept Psychophys* **78**, (2016).

54. Mongillo, G., Barak, O. & Tsodyks, M. SynaptiC Theory of Working Memory. *Science (1979)* **319**, 1543–1546 (2008).

55. Michel, C. M. & Brunet, D. EEG source imaging: A practical review of the analysis steps. *Front Neurol* **10**, (2019).

56. Polich, J. Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology* vol. 118 Preprint at https://doi.org/10.1016/j.clinph.2007.04.019 (2007).

57. Huang, W. J., Chen, W. W. & Zhang, X. The neurophysiology of P 300 - An integrated review. *Eur Rev Med Pharmacol Sci* **19**, 1480–1488 (2015).

58. Berti, S. & Schröger, E. A comparison of auditory and visual distraction effects: Behavioral and event-related indices. *Cognitive Brain Research* **10**, (2001).

59. Ford, J. M. *et al.* The relationship between P300 amplitude and regional gray matter volumes depends upon the attentional system engaged. *Electroencephalogr Clin Neurophysiol* **90**, (1994).

60. Eissa, T. L. & Kilpatrick, Z. P. Learning efficient representations of environmental priors in working memory. *bioRxiv* 2022.07.05.498889 (2022) doi:10.1101/2022.07.05.498889.

61. de Vries, I. E. J., Slagter, H. A. & Olivers, C. N. L. Oscillatory Control over Representational States in Working Memory. *Trends in Cognitive Sciences* vol. 24 Preprint at https://doi.org/10.1016/j.tics.2019.11.006 (2020).

62. Vijayraghavan, S. & Everling, S. Neuromodulation of Persistent Activity and Working Memory Circuitry in Primate Prefrontal Cortex by Muscarinic Receptors. *Frontiers in Neural Circuits* vol. 15 Preprint at https://doi.org/10.3389/fncir.2021.648624 (2021).

63. D'Esposito, M. & Postle, B. R. The cognitive neuroscience of working memory. *Annu Rev Psychol* **66**, 115–142 (2015).

64. Cavanagh, S. E., Towers, J. P., Wallis, J. D., Hunt, L. T. & Kennerley, S. W. Reconciling persistent and dynamic hypotheses of working memory coding in prefrontal cortex. *Nat Commun* **9**, (2018).

65. Berezuk, C., Scott, S. C., Black, S. E. & Zakzanis, K. K. Cognitive reserve, cognition, and real-world functioning in MCI: A systematic review and meta-analysis. *Journal of Clinical and Experimental Neuropsychology* vol. 43 Preprint at https://doi.org/10.1080/13803395.2022.2047160 (2021).

66. Myers, N. E., Chekroud, S. R., Stokes, M. G. & Nobre, A. C. Benefits of flexible prioritization in working memory can arise without costs. *J Exp Psychol Hum Percept Perform* **44**, 398 (2017).

67. Fischer, J. & Whitney, D. Serial dependence in visual perception. *Nature Neuroscience 2014 17:5* **17**, 738–743 (2014).

68. Bays, P. M. & Husain, M. Dynamic shifts of limited working memory resources in human vision. *Science (1979)* **321**, 851–854 (2008).

69. Ma, W. J., Husain, M. & Bays, P. M. Changing concepts of working memory. *Nature Neuroscience 2014 17:3* **17**, 347–356 (2014).

70. Delorme, A. & Makeig, S. EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J Neurosci Methods* **134**, 9–21 (2004).

71. Lopez-Calderon, J. & Luck, S. J. ERPLAB: an open-source toolbox for the analysis of event-related potentials. *Front Hum Neurosci* **8**, (2014).

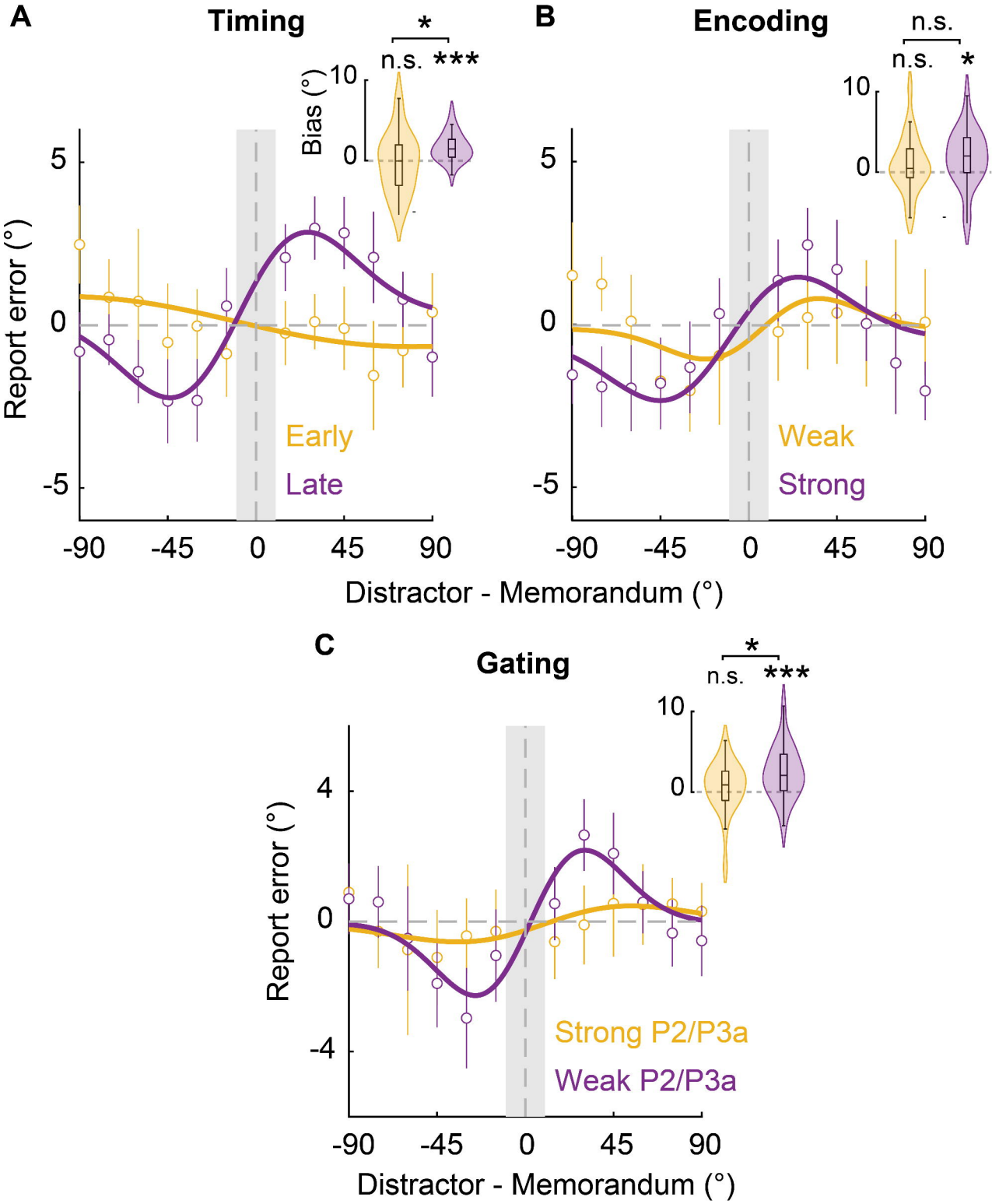
72. Pion-Tonachini, L., Kreutz-Delgado, K. & Makeig, S. ICLabel: An automated electroencephalographic independent component classifier, dataset, and website. *Neuroimage* **198**, 181–197 (2019).

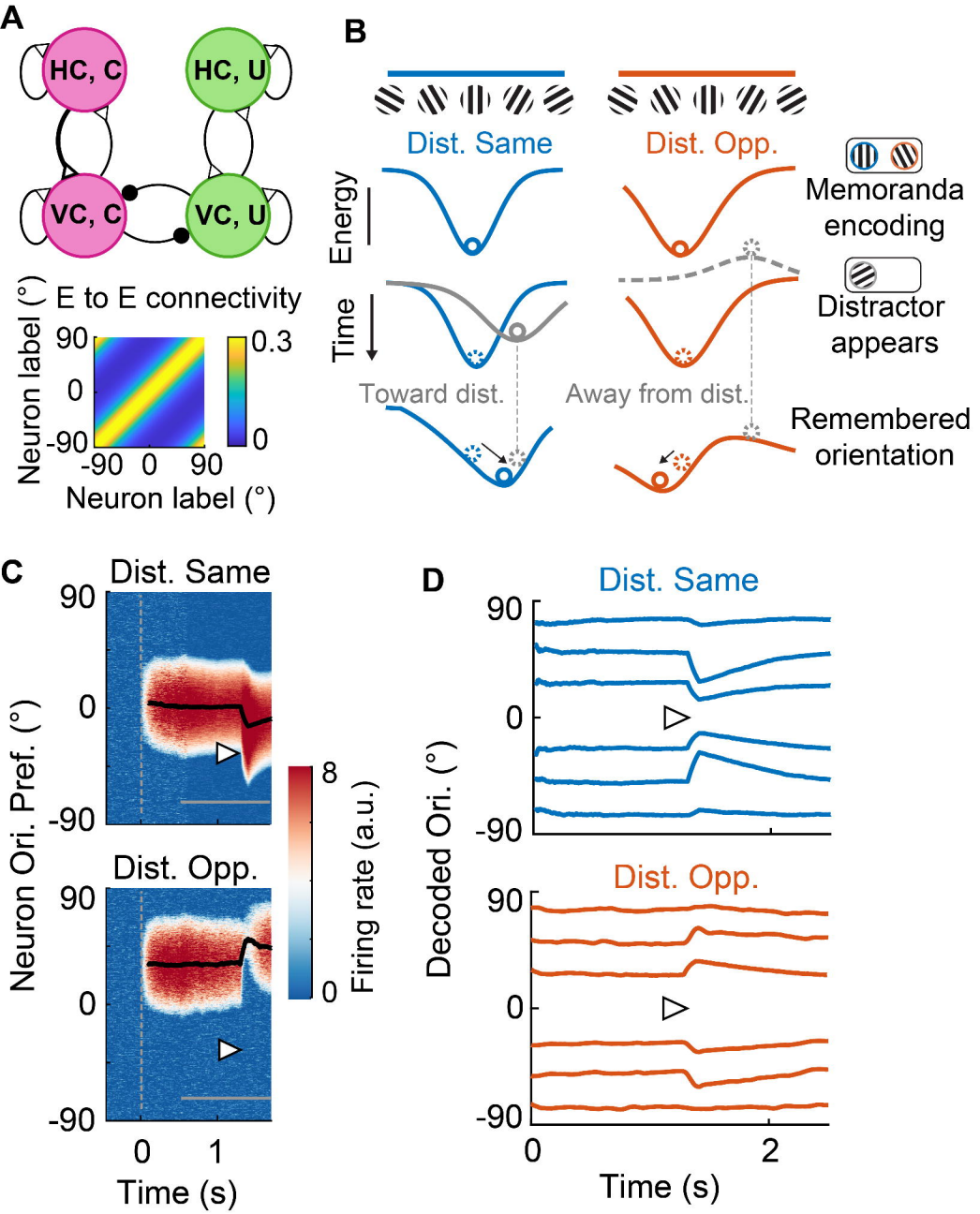
73. Junghöfer, M., Elbert, T., Tucker, D. M. & Rockstroh, B. Statistical control of artifacts in dense array EEG/MEG studies. *Psychophysiology* **37**, 523–532 (2000).

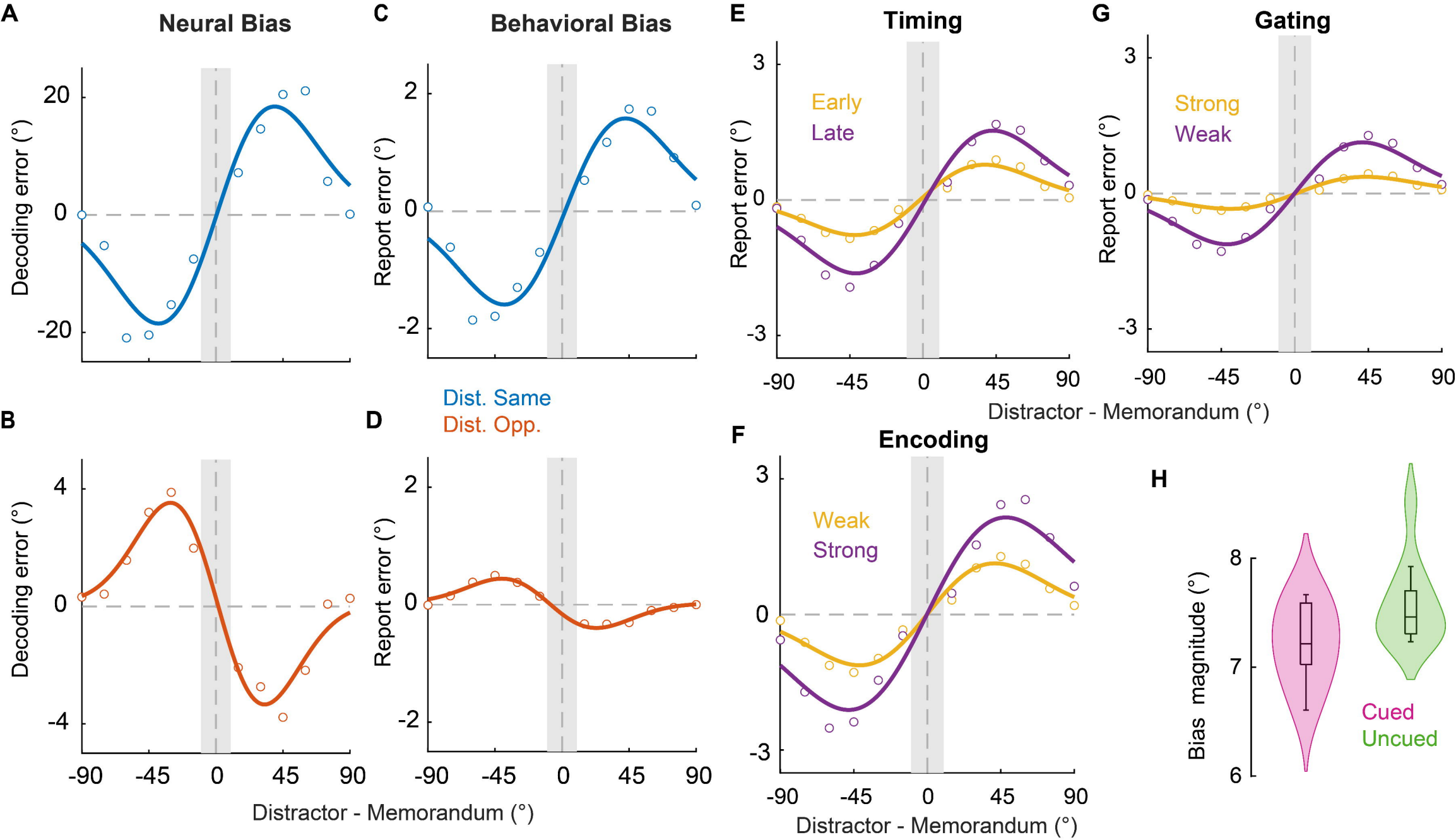
74. Jarosz, A. F. & Wiley, J. What Are the Odds? A Practical Guide to Computing and Reporting Bayes Factors. *J Probl Solving* **7**, 2 (2014).

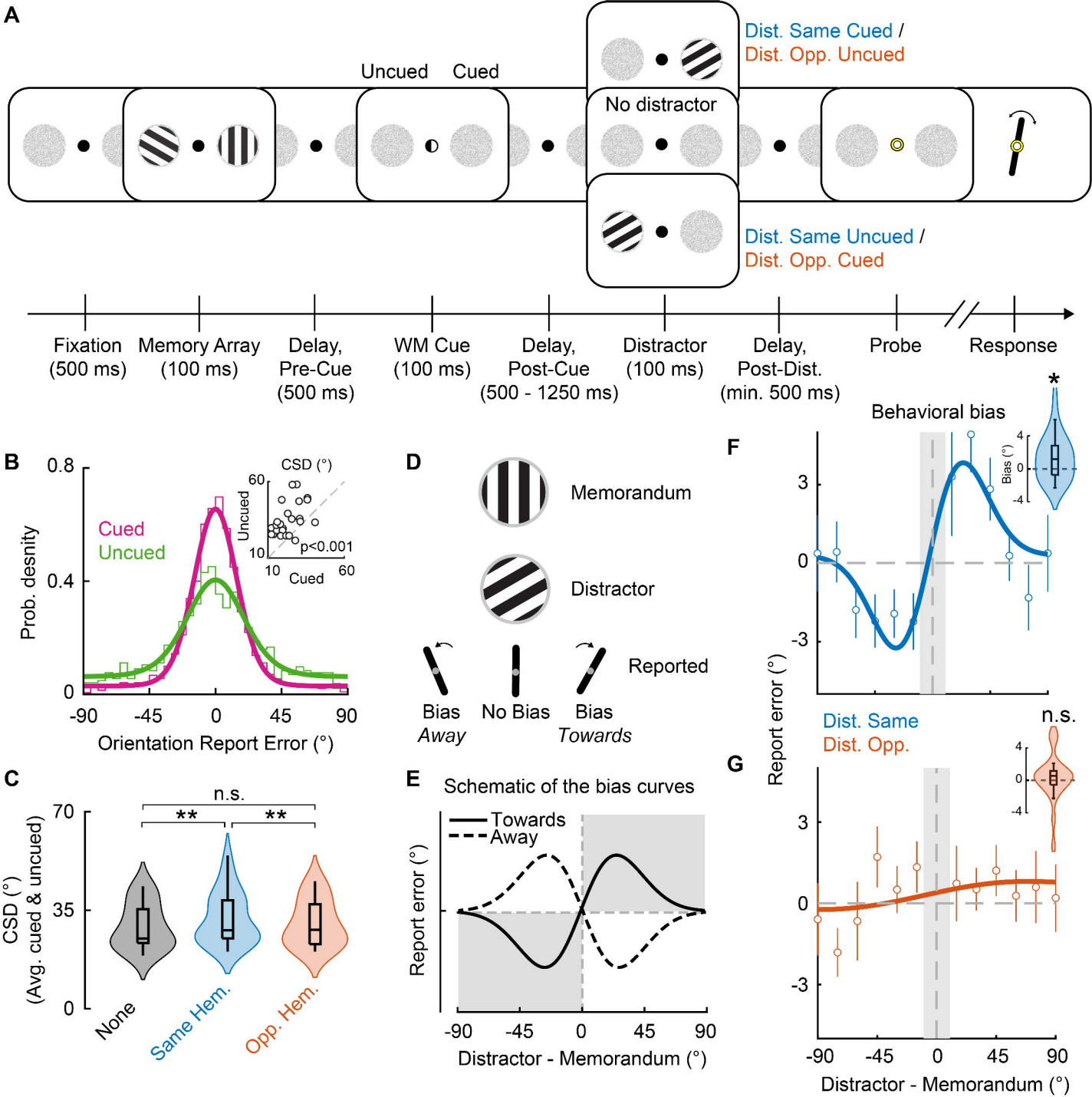
75. Maris, E. & Oostenveld, R. Nonparametric statistical testing of EEG- and MEG-data. *J Neurosci Methods* **164**, 177–190 (2007).

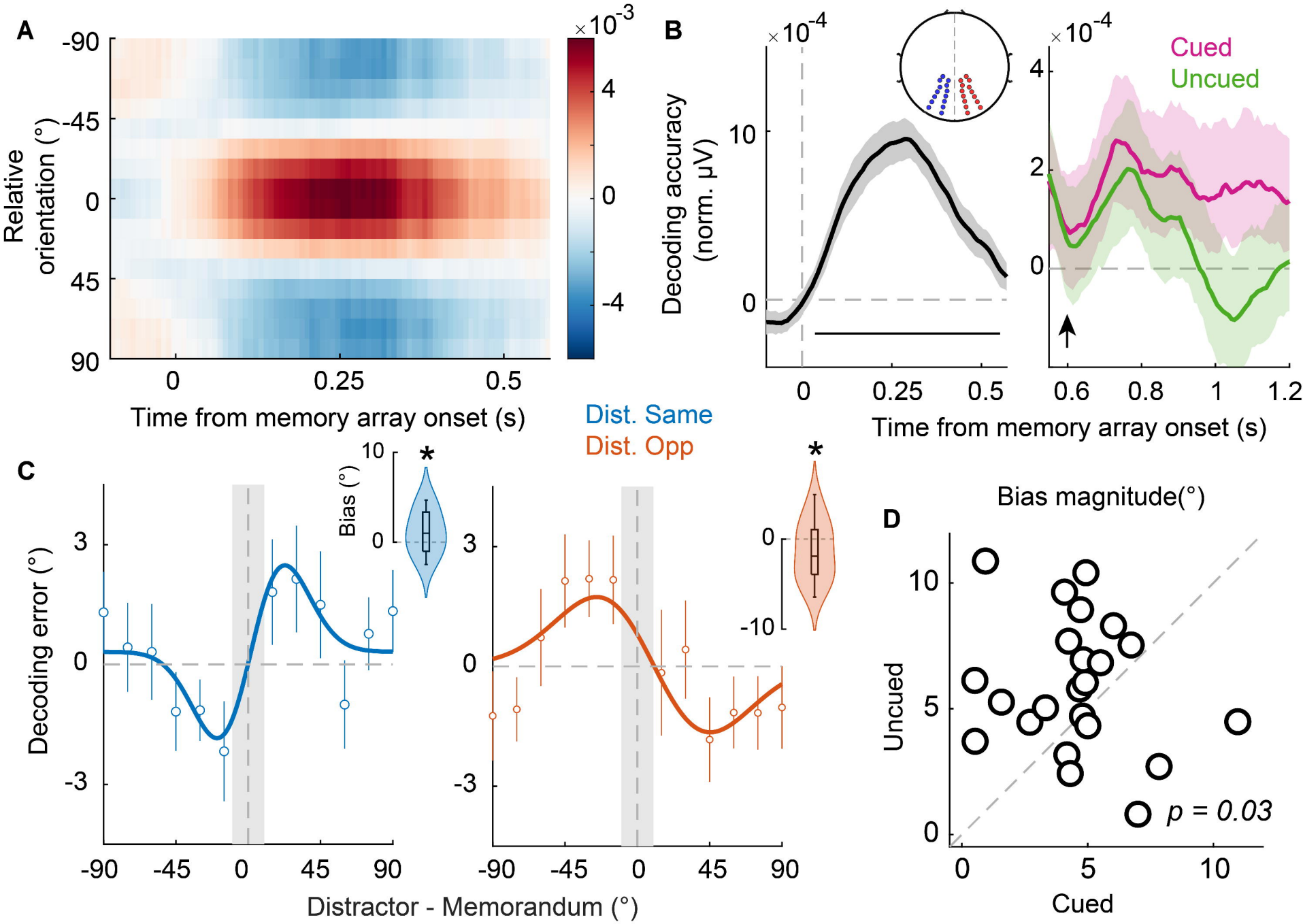
76. Bart Krekelberg. Bayes factor toolbox. Preprint at https://zenodo.org/records/13744717 (2024).

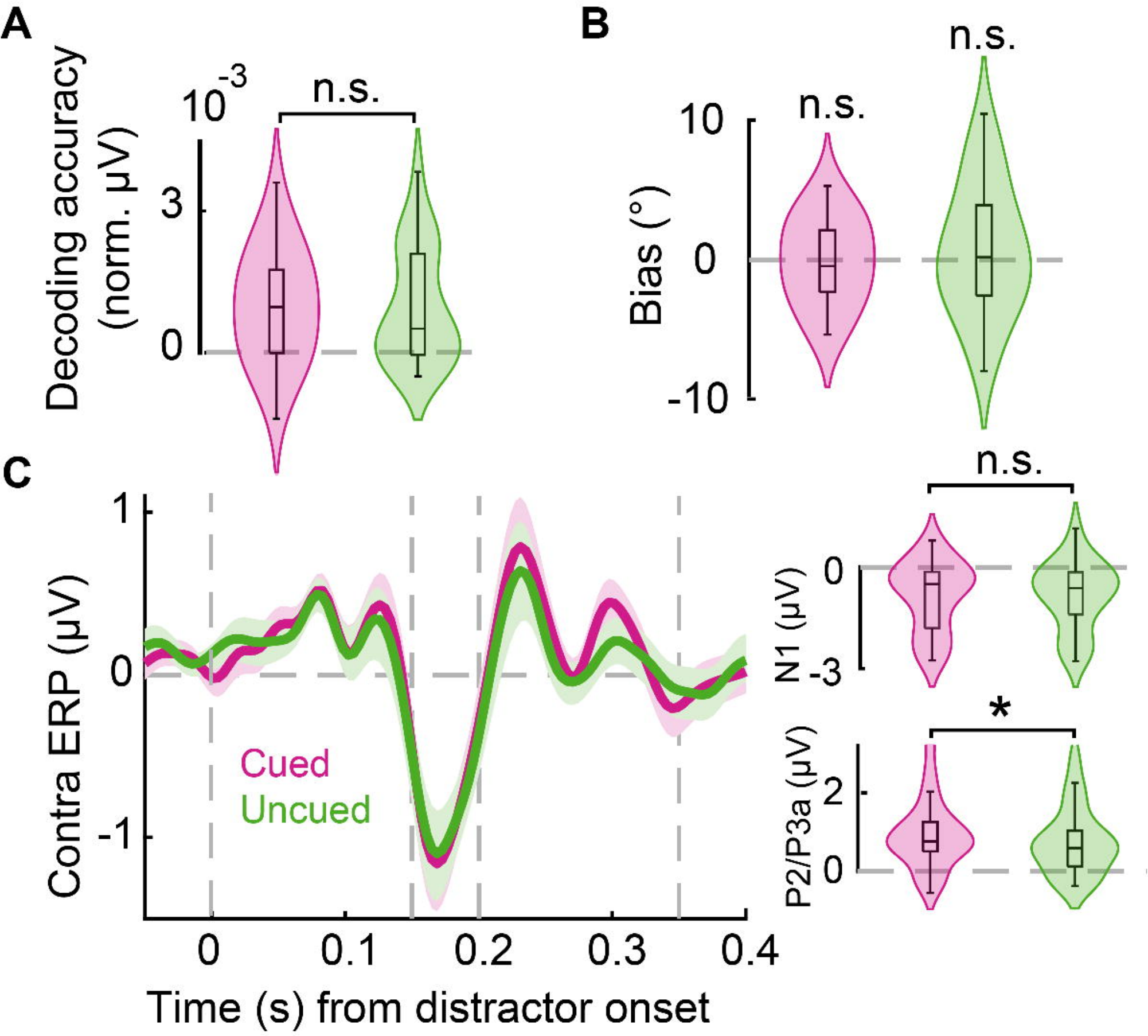

A
Timing
B
Encoding
C
Gating
Report error (°)
Bias (°)
n.s.
*
***
Early
Late
Weak
Strong
Strong P2/P3a
Weak P2/P3a
Distractor - Memorandum (°)


A
HC, C
HC, U
VC, C
VC, U
E to E connectivity
Neuron label (°)
90
0
-90
-90
0
90
Neuron label (°)
0.3
0
B
Dist. Same
Dist. Opp.
Energy
Time
Memoranda encoding
Distractor appears
Toward dist.
Away from dist.
Remembered orientation
C
Dist. Same
Dist. Opp.
Neuron Ori. Pref. (°)
90
0
-90
8
0
Firing rate (a.u.)
0
1
Time (s)
D
Dist. Same
Dist. Opp.
Decoded Ori. (°)
90
0
-90
0
2
Time (s)


A
Neural Bias
Decoding error (°)
20
0
-20
C
Behavioral Bias
Report error (°)
2
0
-2
E
Timing
Early
Late
Report error (°)
3
0
-3
G
Gating
Strong
Weak
Report error (°)
3
0
-3
-90
-45
0
45
90
Distractor - Memorandum (°)
B
Decoding error (°)
4
0
-4
Dist. Same
Dist. Opp.
D
Report error (°)
2
0
-2
F
Encoding
Weak
Strong
3
0
-3
H
Bias magnitude (°)
8
7
6
Cued
Uncued


A
Uncued
Cued
No distractor
Dist. Same Cued /
Dist. Opp. Uncued
Dist. Same Uncued /
Dist. Opp. Cued
Fixation (500 ms)
Memory Array (100 ms)
Delay, Pre-Cue (500 ms)
WM Cue (100 ms)
Delay, Post-Cue (500 - 1250 ms)
Distractor (100 ms)
Delay, Post-Dist. (min. 500 ms)
Probe
Response
B
Cued
Uncued
Prob. desnity
Orientation Report Error (°)
CSD (°)
Uncued
Cued
p<0.001
C
n.s.
**
**
CSD (°)
(Avg. cued & uncued)
None
Same Hem.
Opp. Hem.
D
Memorandum
Distractor
Reported
Bias Away
No Bias
Bias Towards
E
Schematic of the bias curves
Towards
Away
Report error (°)
Distractor - Memorandum (°)
F
Behavioral bias
Bias (°)
Report error (°)
G
Dist. Same
Dist. Opp.
n.s.
Distractor - Memorandum (°)


A
Relative orientation (°)
-90
-45
0
45
90
0
0.25
0.5
Time from memory array onset (s)
×10^-3
4
0
-4
B
Decoding accuracy (norm. µV)
×10^-4
10
0
0
0.25
0.5
Time from memory array onset (s)
×10^-4
4
2
0
Cued
Uncued
0.6
0.8
1
1.2
C
Decoding error (°)
3
0
-3
-90
-45
0
45
90
Distractor - Memorandum (°)
Bias (°)
10
0
*
Dist. Same
Dist. Opp
0
-10
*
D
Bias magnitude(°)
Uncued
Cued
10
5
0
p = 0.03


A
Decoding accuracy (norm. μV)
$10^{-3}$
3
0
n.s.
B
Bias (°)
10
0
-10
n.s.
n.s.
C
Contra ERP (μV)
1
0
-1
Cued
Uncued
0
0.1
0.2
0.3
0.4
Time (s) from distractor onset
N1 (μV)
0
-3
n.s.
P2/P3a (μV)
2
0
*