Changes in behavioral priority influence the accessibility but not the quality of working memory content.

Edward F. Ester[1], Paige Pytel[2]

[1]Department of Psychology and Neuroscience Graduate Program, University of Nevada, Reno
[2]Department of Psychology, Florida Atlantic University

Funding: Supported by NSF 2050833 (E.F.E.)

Author Contributions: E.F.E. conceived and designed the experiment. E.F.E. and P.P. collected the data, analyzed the data, and co- wrote the manuscript.

Data & Code Availability: Stimulus presentation software, raw and preprocessed data, and analytic software associated with this project are freely available at https://osf.io/rg8h4/

Abstract

Efficient behavioral selection requires rapid comparisons of sensory inputs with internal representations of motor affordances and goal states, and many of these comparisons take place in working memory (WM). Selecting an item stored in WM is known to blunt and/or reverse information loss in stimulus-specific representations of that item reconstructed from human brain activity, but extant studies have focused on all-or-none circumstances that allow or prevent an agent from selecting one item stored in WM. Behavioral studies suggest that humans can flexibly assign different levels of priority to different items stored in WM, but how doing so influences neural representations of WM content is unclear. One possibility is that assigning different levels of priority to items in memory influences the quality of those representations, resulting in more robust neural representations of high- vs. low-priority WM content. A second (non-exclusive) possibility is that asymmetries in behavioral priority influence how rapidly neural representations of high- vs. low-priority WM content can be selected and reported. To test these alternatives, we decoded high- and low-priority WM content from EEG recordings obtained while human volunteers (both sexes) performed a retrospectively cued WM task. Probabilistic changes in the behavioral relevance of a remembered item had no effect on asymptotic decoding performance; instead, these changes influenced the latency at which peak decoding performance was reached. Thus, our results indicate that probabilistic changes in the behavioral relevance of WM content influence the ease with which memories can be accessed and retrieved independently of their strength.

Flexible behaviors require sensory inputs to be compared with internal representations of goal states. Many of these comparisons take place in working memory (WM), a capacity- and duration-limited system that forms a temporal bridge between fleeting sensory phenomena and possible actions (van Ede & Nobre, 2022). Changing environmental circumstances and evolving behavioral goals necessitate the existence of internal selection mechanisms that prioritize task-relevant WM contents for action, especially when an agent must select from among multiple prospective actions or execute a series of actions in sequence (e.g., Cisek & Kalaska, 2010; Cisek, 2019). For example, making your favorite cup of coffee involves performing a series of actions in a specific order, and only when certain external conditions – e.g., the water in the kettle has begun to boil – are met.

The neural consequences of assigning priority to specific WM content can be studied by measuring brain activity linked with WM storage while participants perform a retrospectively cued memory task (Griffin & Nobre, 2003; Lewis-Peacock et al., 2012; Ester et al., 2018; Panichello & Buschman, 2021). In a typical retrocue experiment, an agent remembers an array of items over a brief delay and use this information to perform a memory-guided behavior. During storage, an informative cue instructs the observer which remembered item(s) are most likely to be required for action at the end of the trial. The typical finding is that an informative retrocue improves memory performance relative to an uninformative cue or no-cue condition (see Souza & Oberauer, 2016, and Myers et al., 2017, for recent comprehensive reviews). Moreover, improvements in memory performance are typically accompanied by improvements in the quality of stimulus-specific WM representations reconstructed from human brain

activity, with informative retrocues arresting or even reversing information loss that accumulates during WM storage in the absence of a cue (Sprague et al., 2014; Sprague et al., 2016; Ester et al., 2018; Nouri & Ester, 2020).

With few exceptions (Berryhill et al., 2012; Shimi et al., 2014; Günseli et al., 2015; Günseli et al., 2019) most retrocue studies have used perfectly reliable cues. That is, when an informative retrocue appears, it indicates which of a set of remembered items will be later probed with perfect accuracy. Living organisms, however, exist in dynamic natural environments where the future can take on several possibilities. Thus, the likelihood that that any one piece of information stored in WM will become behaviorally relevant is best understood as a matter of probability rather than a certainty. Behavioral studies suggest that retrocue benefits on WM performance are probabilistic and scale with cue reliability; thus, human observers can flexibly assign different levels of behavioral priority to different items stored in memory. However, less is known about how graded changes in behavioral relevance influence neural representations of WM content. One possibility is that probabilistic changes in behavioral relevance modulate memory strength, for example by facilitating the allocation of attentional gain to neural populations encoding high- vs. low-priority WM content (e.g., Bays & Taylor, 2017). Conversely, a second – and non-exclusive – possibility is that probabilistic changes in behavioral relevance influence modulate how easily memories can be selected for behavioral read-out (e.g., e.g., Souza et al., 2016). A critical test of these alternatives would offer new insights into how internal selective attention is used to flexibly prioritize task-relevant WM content. Here, we provide this test.

In two experiments, we recorded EEG while human volunteers performed a retrospectively cued spatial WM task. Across experimental blocks, we varied retrocue reliability between 100% (i.e., perfectly predictive) or 75% and quantified how this manipulation influenced our ability to decode remembered information from scalp EEG measurements. Cue reliability had no influence on maximum decoding performance in either experiment; instead, it had a large effect on the latency at which maximum decoding performance was reached. Thus, we show that probabilistic changes in behavioral priority affect the accessibility but not the quality of stimulus-specific WM representations.

Methods

Participants. A total of 71 human volunteers (both sexes) were enrolled in this study. 36 participants completed Experiment 1 and 35 participants completed Experiment 2. Participants were recruited from the Florida Atlantic University community and completed a single 2.5-hour testing session in exchange for monetary remuneration ($15/h in amazon.com gift cards). All participants self-reported normal or corrected-to-normal visual acuity and color vision. All study procedures were approved by the local institutional review board, and all participants gave both written and oral informed consent prior to enrolling in the study. Two participants in Experiment 1 and two participants in Experiment 2 voluntarily withdrew from the study prior to completing both cue conditions (i.e., 100% valid vs. 75% valid); data from these participants were excluded from final analyses. Thus, the data reported here reflect 34 participants in Experiment 1 and 33 participants in Experiment 2.

Testing Environment. Participants were seated in a dimly-lit and acoustically shielded recording chamber for the duration of the experiment. Stimuli were generated in MATLAB and rendered on a 17'' Dell CRT monitor cycling at 85 Hz (1024 x 768 pixel resolution) via PsychToolbox-3 software extensions (Kleiner et al., 2007). Participants were seated approximately 65 cm from the display (head position was unconstrained). To combat fatigue and/or boredom, participants were offered short breaks at the end of each experimental block.

Experiment 1 - Spatial Postcue Task. A task schematic is shown in Figure 1. Each trial began with an encoding display lasting 500 ms. The encoding display contained two colored circles (blue and red; subtending 1.75 degrees visual angle [DVA] from a

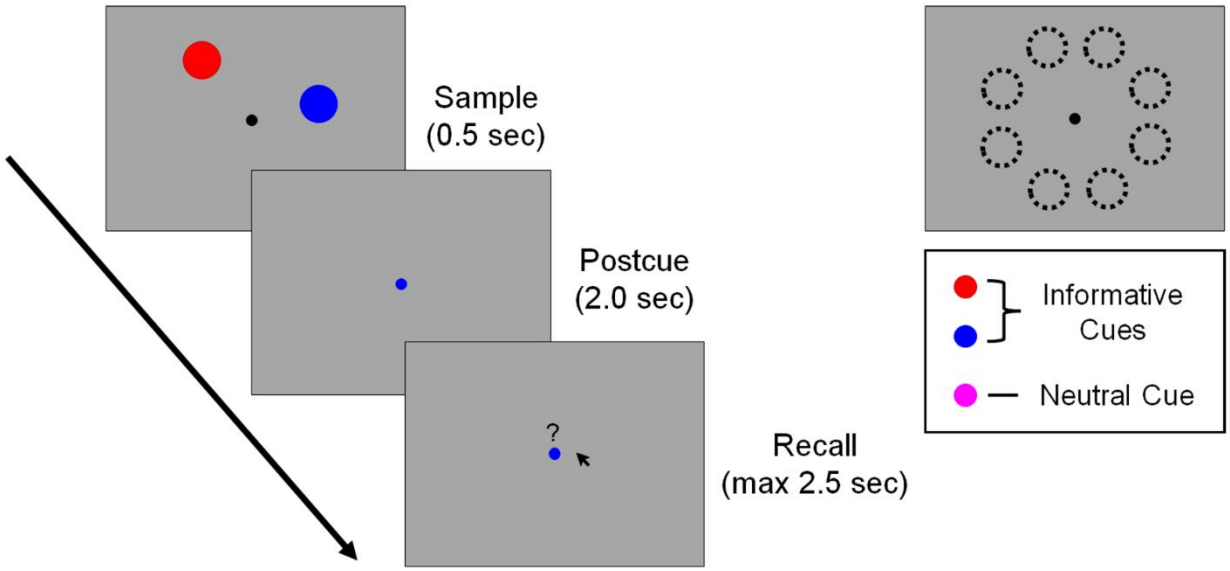

Figure 1. Spatial Postcue Task. Participants encoded the positions of two discs into memory. A postcue instructed participants to remember one disc (informative trials) or both discs (neutral trials) over a blank delay. Participants then recalled the position of the disc matching the color of the fixation point. Discs always appeared in two of eight possible positions (upper right), with disc positions counterbalanced across the experiment. During the first half of experimental blocks postcue reliability was fixed at 100%; during the second half of experimental blocks reliability was lowered to 75%. See text for additional details. Note: the displays shown here are included for exposition and not drawn to scale; see Methods for detailed information regarding stimulus and display geometry.

viewing distance of 65 cm) rendered at of eight polar locations (22.5° to 337.5° in 45° increments) along the perimeter of an imaginary circle (radius 7.5 DVA) centered on a circular fixation point (subtending 0.25 DVA) rendered in the middle of the display. The locations of the two discs were counterbalanced across each task (i.e., 100% valid vs. 75% valid), though not necessarily within an experimental block. Participants were instructed to maintain fixation and refrain from blinking for the duration of each trial. The encoding display was followed by a 2.0 sec postcue display. During informative cue trials the fixation point changed colors from black to either blue or red (i.e., matching the color of a remembered disc), while during neutral trials the fixation point changed colors from black to purple (e.g., the “average” of blue and red). The postcue display was

followed by a probe display containing a blue or red fixation point, mouse cursor, and question mark symbol (the initial position of the mouse cursor was always on top of the fixation point). Participants were required to click on the location of the disc matching the color of the fixation point. Participants were instructed to prioritize accuracy over speed, but a 2.5 sec response deadline was imposed. The trial ended when participants clicked on a location or the response deadline elapsed. Sequential trials were separated by a 1.5-2.5 sec blank period (randomly and independently selected from a uniform distribution after each trial).

Each participant completed 12-18 blocks of 56 trials in this task. During the first half of experimental blocks postcue reliability was fixed at 100%, i.e., during informative trials the postcue perfectly predicted which disc would be probed at the end of the trial. During the second half of experimental blocks postcue reliability was reduced to 75%. During valid cue trials the fixation point remained the same color during the postcue and probe displays; during invalid cue trials the fixation point changed from one color during the postcue display (e.g., blue) to a different color during the probe display (e.g., red). In both cases, participants were instructed to click on the location of the disc matching the color of the fixation point during the probe period.

Experiment 2 – Spatial Retrocue Task. Experiment 2 was identical to Experiment 1, with the exception that the postcue was replaced by a retrocue presented midway through the delay period (i.e., 1.0 sec after termination of the sample display). We use the term postcue in Experiment 1 and retrocue in Experiment 2 to emphasize that in the former experiment participants can consult an iconic representation of disc positions while in the latter experiment they cannot (Souza & Oberauer, 2016).

EEG Acquisition and Preprocessing. Continuous EEG was recorded from 63 uniformly distributed scalp electrodes using a BrainProducts “actiCHamp” system. The horizontal and vertical electrooculogram (EOG) were recorded from bipolar electrode montages placed over the left and right canthi and above and below the right eye, respectively. Online EEG and EOG recordings were referenced to a 64th electrode placed over the right mastoid and digitized at 1 kHz. All data were later re-referenced to the algebraic mean of the left- and right mastoids, with 10-20 site TP9 serving as the left mastoid reference. Data preprocessing was carried out via EEGLAB software extensions (Delorme & Makeig, 2004) and custom software. Data preprocessing steps included the following, in order: (1) resampling (from 1 kHz to 250 Hz), (2) bandpass filtering (1 to 50 Hz; zero-phase forward- and reverse finite impulse response filters as implemented by EEGLAB), (3) epoching from -1.0 to +5.0 relative to the start of each trial, (4) identification, removal, and interpolation of noisy electrodes via EEGLAB software extensions, and (5) identification and removal of oculomotor artifacts via independent components analysis via EEGLAB. After preprocessing, our analyses focused exclusively on the following 10-20 occipitoparietal electrodes: P7, P5, P3, Pz, P2, P4, P6, P8, PO7, PO3, POz, PO2, PO4, PO8, O1, O2, Oz.

Data Cleanup. Prior to analyzing participants’ behavioral or EEG data, we excluded all trials where the participant responded with a latency of < 0.3 sec (we attributed these trials to accidental mouse clicks following the onset of the probe display rather than a deliberate recall of a specific stimulus position) or failed to respond within the 2.5 sec deadline. This resulted in an average loss (±1 S.E.M.) of 0.429 ± 0.09% trials in Experiment 1 and 0.841 ± 0.22% of trials in Experiment 2.

Decoding Spatial Positions from Posterior Alpha-Band EEG Signals. Location decoding was based on the multivariate distance between EEG activity patterns associated with memory for specific positions. This approach is an extension of earlier parametric decoding methods (Wolff et al., 2017) designed for use in circular feature spaces (Fig 2). Following earlier work (e.g., Ester et al., 2018), we extracted spatiotemporal patterns of alpha-band activity (8-13 Hz) from 17 occipitoparietal electrode sites (see *EEG Acquisition and Preprocessing* above). The raw timeseries at each electrode was bandpass filtered from 8-13 Hz (zero-phase forward-and-reverse filters as implemented by EEGLAB software), yielding a real-valued signal f(t). The analytic representation of f(t) was obtained via Hilbert transformation:

$$z(t) = f(t) + if(t)$$

where *i* is the imaginary operator and *if(t)* = $A(t)e^{i\varphi(t)}$. Alpha power was computed by extracting and squaring the instantaneous amplitude A(t) of the analytic signal z(t).

Location decoding performance was computed separately for each disc (i.e., blue vs. red), trial type (i.e., informative vs. neutral) and each task (i.e., 100% vs. 75%) on a timepoint-by-timepoint basis. In the first phase of the analysis, we sorted data from each condition into 5 unique training and test data sets using stratified sampling while

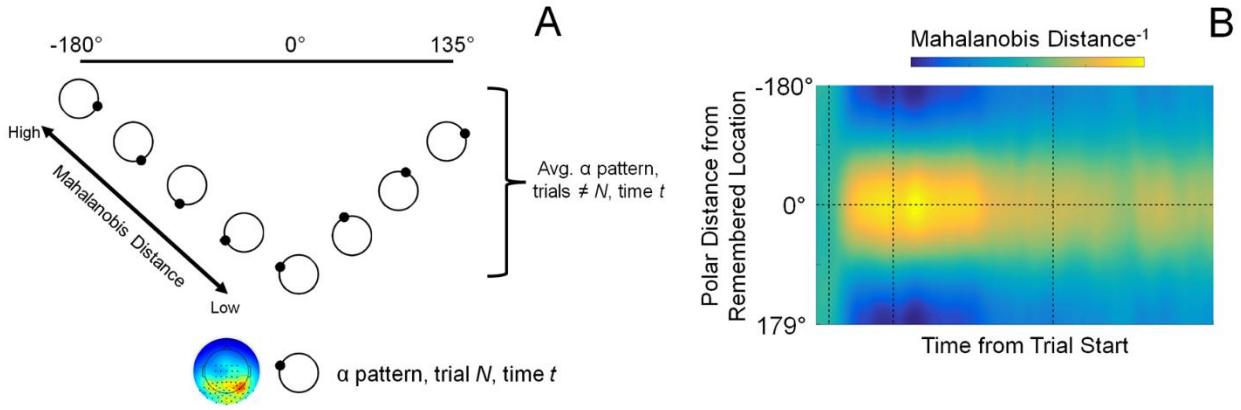

Figure 2. Decoding Analysis. (A) We computed the Mahalanobis distance between patterns of alpha-band activity measured during each trial of a training data set with position-averaged patterns measured in an independent training data set. (B) If patterns of alpha-band activity contain robust position-specific signals, then multivariate distances should be smallest when comparing activation patterns associated with memory for a specific location in the test data set with activation patterns associated with memory for the same location in the training data set, and the multivariate distance between test and training data patterns should increase with polar distance.

ensuring that each training set was balanced across remembered positions (i.e., we ensured that each training data set contained an equal number of observations where the location of the remembered stimulus was at 22.5°, 67.5°, etc.). We circularly shifted the data in each training and test data set to a common center (0°, by convention) and computed trial-averaged patterns of responses associated with memory for each disc position in each training data set. Next, we computed the Mahalanobis distance between trial-wise activation patterns in each test data set with position-specific activation patterns in the corresponding test data set, yielding a location-wise set of distance estimates (Fig 2A). If scalp activation patterns contain information about remembered positions then distance estimates should be smallest when comparing patterns associated with memory for identical positions in the training and test data set and largest when comparing opposite positions (i.e., those ±180° apart), yielding an inverted gaussian-shaped function (e.g., Fig 2B). Trial-wise distance functions were

averaged, sign-reversed, and convolved with a cosine function to yield a single decoding estimate for condition and time point with chance decoding yielding a value of 0. Decoding results from each training- and test-data set pair were averaged (thus ensuring the internal reliability of our approach), yielding a single decoding estimate per participant, timepoint, and task condition.

Quantifying Peak Decoding Accuracy. To determine whether changes in cue reliability influenced the strength of location-specific representations stored in WM, we calculated peak decoding accuracy during 100% reliable and 75% reliable blocks. Peak decoding estimates were then compared with a bootstrap test. We first selected (with replacement) and averaged time courses of decoding activity for the probed location from *N* of *N* participants. Next, we calculated the time of maximum decoding performance following the onset of the probe display (i.e., 2.5 sec after the start of each trial). We defined a 100 ms window around peak decoding performance (i.e., 50 ms before peak decoding performance to 50 ms after it) and used this window to compute temporally averaged peak decoding performance in the same sample of *N* participants. These calculations were performed separately for data from 100% valid and 75% valid blocks and permuted 10,000 times, with a new subsample of participants chosen on every permutation. Finally, we computed the average and 95% confidence interval of peak decoding performance during 100% valid and 75% valid blocks. Statistical significance was assessed by counting the proportion of permutations where peak decoding performance was larger during 100% valid blocks compared to 75% valid blocks.

Quantifying Peak Decoding Latency. To determine whether changes in cue reliability influenced the timing of WM retrieval, we compared the latencies of above-chance decoding performance during the cue and probe displays. During neutral trials, we computed a cross-correlation between average probe-matching decoding performance from 0.0 to 1.5 sec after the onset of the probe display. Specifically, we calculated the normalized correlation coefficient between the time course of decoding performance during the 100% and 75% blocks while temporally shifting the latter relative to the former by -1.0 to +1.0 sec in 4 ms increments. If time courses of decoding performance during 100% blocks and 75% blocks are identical, then the maximum cross-correlation should be observed at a temporal lag of 0.0 sec. Conversely, if the time-course of above-chance decoding performance during 100% valid blocks precedes the time-course of above-chance decoding performance during 75% valid blocks, then the maximum cross-correlation should occur at a negative temporal lag (i.e., when the time course of decoding performance during 75% blocks is shifted earlier in time). We compared the observed cross correlation function with a null distribution of cross-correlation function estimated by shuffling participant level condition labels (i.e., 100% valid vs. 75% valid) 10,000 times. An analogous approach was used to quantify temporal lags in decoding performance during informative cue trials (whether valid or invalid), with the exception that we used a window spanning -0.5 to 1.5 sec relative to the onset of the cue display. A broader window was deliberately selected as we anticipated cue-matching decoding performance to exceed chance levels during the post- or retrocue period when an informative cue was present.

Statistical Comparisons – Behavioral Data. Participants' behavioral data (i.e., absolute average recall error and average response time) were analyzed using standard repeated-measures parametric statistics (e.g., t-test, ANOVA); for these comparisons we report test statistics, p-values, and effect size estimates.

Statistical Comparisons – EEG Data. The decoding analysis we used assumes chance-level decoding performance of 0. Likewise, direct comparisons of decoding performance or reconstruction strength across conditions (e.g., 100% vs. 75%) assume a null statistic of 0. Thus, we evaluated decoding performance by generating null distributions of decoding performance (or differences in decoding performance across conditions) by randomly inverting the sign of each participant's data with 50% probability and averaging the data across participants. This procedure was repeated 10,000 times, yielding a 10,000-element null distribution for each time point. Finally, we implemented a cluster-based permutation test (Maris & Oostenveld, 2007; Wolff et al. 2017) with cluster-forming and cluster-size thresholds of $p < 0.05$ to correct for multiple comparisons across time points. Differences in peak decoding accuracy were quantified with bootstrap tests (see *Quantifying Peak Decoding Accuracy)*, and differences in decoding latency were quantified via randomization tests (see *Quantifying Peak Decoding Latency*)

Results

Experiment 1 - Postcues

Behavioral Performance. Participants' memory performance was quantified via average absolute recall error (i.e., the angular difference between the polar location of the probed stimulus and the polar location reported by the participant) and average response latency. During 100% blocks, participants received a neutral or perfectly informative postcue. Conversely, during 75% blocks participants received a neutral, valid, or invalid postcue. Thus, we initially analyzed data from 100% and 75% blocks separately. Participants' recall errors during 100% blocks and 75% blocks are summarized in Figure 3A. During 100% blocks, participants recall errors were significantly lower during valid relative to neutral trials [$t(33) = 4.178$, $p < 0.0002$, $d = 0.474$]. During 75% blocks, a repeated-measures analysis of variance (ANOVA) with cue type (neutral, valid, or invalid) as the sole factor revealed a main effect, [$F(2,66) = 4.456$, $p = 0.015$, $\eta^2 = 0.119$]. To facilitate performance comparisons across different levels of cue reliability (i.e., 100% vs. 75%) we calculated within-condition cue effects by (a) computing differences in recall error during 100% valid and 100% neutral trials and (b) computing differences in recall error between 75% valid, 75% neutral, and 75% invalid trials (Figure 3B). Direct comparisons of cue effects during 100% valid and 75% valid blocks indicated that valid cues lowered recall errors by equal amounts during 100% and 75% valid blocks [$M = -0.699°$ and $-0.706°$, respectively, $t(33) = 0.038$, $p = 0.969$]. Conversely, invalid cues during 75% blocks did not incur a performance cost; if anything, recall errors were marginally lower during invalid relative to neutral trials [$M = 9.183°$ vs. $9.225°$; $t(33) = 0.166$, $p = 0.869$].

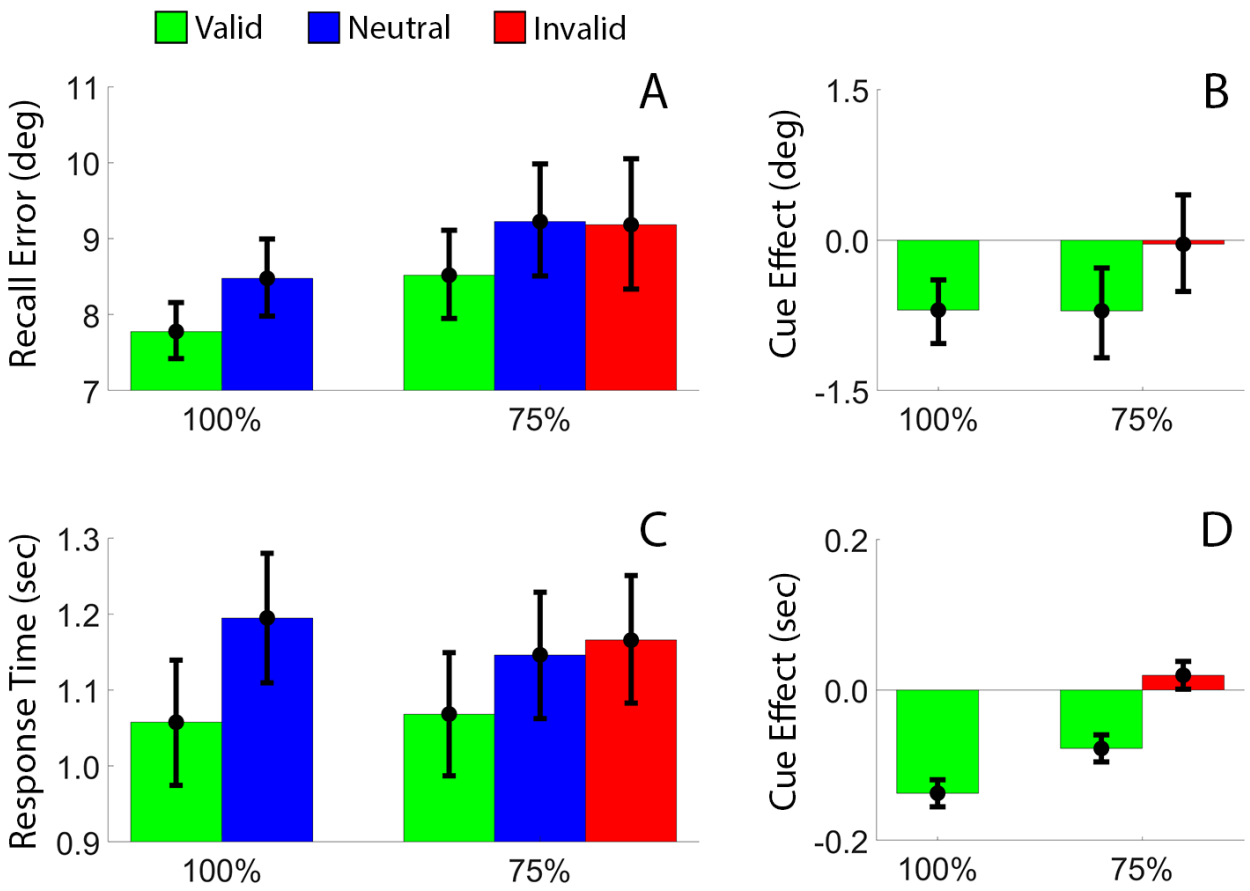

Figure 3. Memory Performance in Experiment 1. (A) Average absolute recall error as a function of cue type (valid, neutral, invalid) and cue reliability (100%, 75%). (B) Cue effects, defined as the difference between valid/invalid and neutral trials. (C) Average response times and (D) cue effects. Error bars depict the 95% confidence interval of the mean.

A complementary analysis of participants' response times (Figure 3C) revealed a facilitatory effect of valid vs. neutral cues during 100% blocks [$t(33) = 14.540$, $p < 1e\text{-}16$, $d = 0.543$], and a repeated measures ANOVA revealed a significant effect of cue type during 75% blocks [$F(2,66) = 46.749$, $p < 1e\text{-}14$; $\eta^2 = 0.586$]. Analyses of cue effects (Figure 3D) revealed a significantly larger effect of valid cues on response times during 100% vs. 75% blocks [M = 0.136 vs. 0.078 sec, respectively; $t(33) = 5.884$, $p < 1.35e\text{-}6$, $d = 1.088$]. Once again, invalid cues during 75% blocks did not incur a significant performance cost (M = 19 ms; $t(33) = 2.006$, $p = 0.0532$, $d = 0.078$). Thus, valid cues

presented during 100% and 75% blocks led to equal improvements in recall performance and parametric improvements in response times compared to neutral cues. Conversely, invalid cues did not significantly impair either recall performance or response times.

EEG Decoding Performance. We used a decoding analysis (Figure 2) to quantify how changes in behavioral priority influenced location-specific representations stored in WM. During neutral trials, an uninformative postcue instructed participants to remember the positions of both discs presented in the sample display; 2.0 sec later a probe display prompted participants to report the location of one disc via a mouse click. Based on earlier findings (e.g., Ester et al., 2018; Nouri & Ester, 2019; Ester & Nouri, 2022) expected equivalent decoding performance for each disc during the sample and postcue displays, but significantly higher decoding performance for the probed relative to unprobed disc during the probe display. Moreover, we expected equivalent performance across different levels of cue reliability, i.e., 100% vs. 75% blocks. These predictions were borne out in analyses of location decoding performance during neutral trials (Figure 4). Decoding performance for the disc that was ultimately probed and the disc that was not ultimately probed increased rapidly during the sample display but fell back to chance levels during the ensuring postcue display. Decoding performance for the probe-matching disc – but not the probe-nonmatching disc – increased rapidly following onset of the probe display before returning to chance levels shortly after participants responded. Nearly identical patterns of decoding performance were

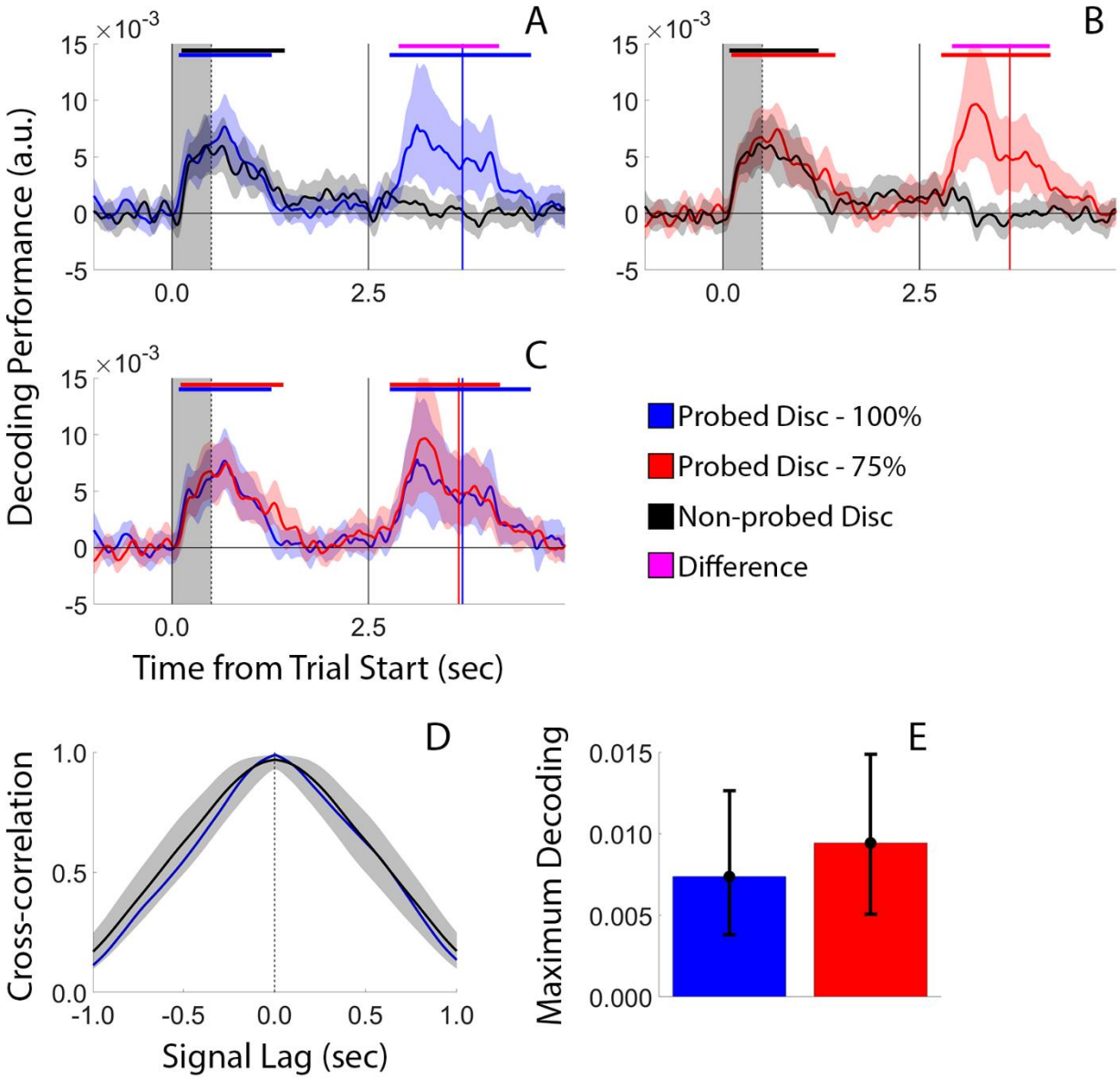

Figure 4. Decoding Performance During Neutral Trials. (A) Average decoding performance for the probe-matching and probe-nonmatching discs during neutral trials of 100% reliable blocks. (B) Identical to (A), but for neutral trials of 75% blocks. (C) Overlay of probe-matching decoding performance from 100% reliable and 75% blocks. (D) Cross-correlation analysis between time-courses of probe-locked decoding performance. (E) Average peak decoding accuracy during 100% and 75% blocks. Shaded area in panels (A-C) mark the sample display; vertical lines at time 0.0, 0.5, and 2.5 mark the onset of the sample, postcue, and response displays, respectively. Blue and red vertical lines mark the average response time across participants. Horizontal bars at the top of each plot mark epochs where decoding performance was significantly > 0 or epochs where decoding performance was significantly greater for the probe-matching vs. non-matching stimulus. Shaded regions and errorbars depict the 95% confidence interval of the mean.

observed during 100% valid blocks (Figure 4A) and 75% valid blocks (Figure 4B), and

cue reliability had no differences in the peak magnitude or latency of probe-matching decoding performance (Figures 4C-E). These findings were expected and are a straightforward replication of earlier results (Ester et al., 2018; Nouri & Ester, 2020).

Next, we examined location decoding performance during informative cue trials (Figure 5). During 100% reliable blocks, valid cues informed participants which disc would be probed with complete certainty. This condition is a direct replication of our earlier work (Ester et al., 2018) in which we found that valid postcues slowed or presented the gradual return to chance-level decoding performance seen during neutral trials. The current data replicate this finding (Figure 5A): during 100% valid trials decoding performance for the cue-matching disc remained at above-chance levels during the postcue display and into the probe display, while decoding performance for the cue-nonmatching disc quickly returned to chance levels following the appearance of the postcue. A qualitatively different pattern emerged during 75% reliable blocks (Figure 5B). Since invalid cues had no deleterious effect on participants' recall errors (Figure 3B) or response times (Figure 3D), our analyses of decoding performance pooled across valid and invalid cue trials. Decoding performance for the cue-matching and cue-nonmatching discs returned to chance levels during postcue period, while decoding performance for the probe-matching disc (whether a valid or invalid trial) increased rapidly after the appearance of the probe display and remained at above-chance levels until after participants responded. Direct comparisons of cue- and probe-matching decoding performance during 100% and 75% blocks (Figure 5C) revealed that maximum decoding performance was reached significantly earlier during 100% blocks than 75% blocks (Figure 5D), even though averaged peak decoding performance was

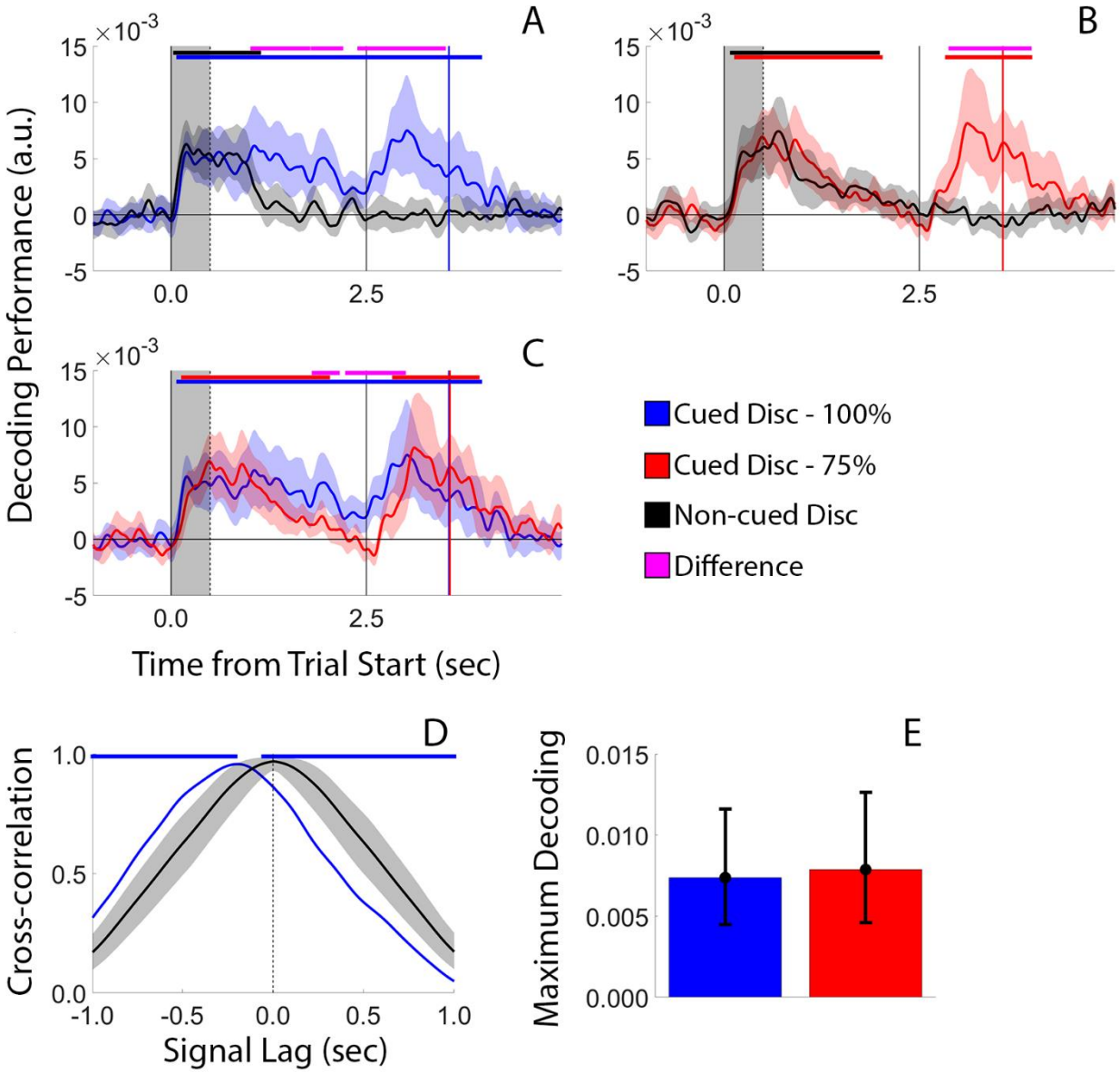

Figure 5. Decoding Performance During Informative Cue Trials. Conventions are identical to Figure 4.

identical across these conditions. Thus, the results of this analysis support the hypothesis that changes in behavioral priority affect the accessibility but not the strength of WM representations.

Experiment 2 – Retrocues

Following earlier work (e.g., Souza & Oberauer, 2016) we use the term postcue to refer to any event informing which of a set of remembered item(s) will be tested that occurs immediately after encoding (including instances where an agent can apply this information to stimulus representations in sensory memory) and retrocue to refer to any informative event occurring after memory consolidation is complete. Prior evidence suggests that post- and retrocues engage separate but complementary mechanisms that promote storage of high-quality neural representations of memoranda (Ester et al., 2018). Thus, Experiment 2 was conducted to replicate and extend the findings of Experiment 1 and earlier studies. Experiment 2 was identical to Experiment 1, with the exception that postcues were replaced with retrocues presented midway through the blank interval separating the sample and probe displays (i.e., 1.0 sec after termination of the sample display). Once again, cue reliability was varied between 100% and 75% in different experimental blocks.

Behavioral Performance. Behavioral data from Experiment 2 were analyzed identically to Experiment 1. Participants' recall errors during 100% blocks and 75% blocks are summarized in Figure 6A. Recall errors were significantly lower during valid compared to neutral trials during 100% blocks [$t(32) = 2.637$, $p = 0.013$, $d = 0.179$]. A repeated-measures analysis of variance (ANOVA) applied to recall errors during 75% blocks revealed a main effect of cue type (i.e., valid, invalid, neutral), [$F(2,64) = 8.066$, $p = 0.0008$, $\eta^2 = 0.201$]. Cue effects - defined as the difference in recall errors between valid vs. neutral cues during 100% blocks and the difference in recall errors between valid and invalid vs. neutral cues during 75% blocks - are summarized in Figure 1C. Valid

cues lowered recall errors by an equal amount during 100% and 75% valid blocks [M = 0.335° and 0.814°, respectively, $t(32) = 1.918$, $p = 0.064$], while invalid cues during 75% blocks incurred a significant performance cost compared to neutral trials [M = 1.086°, $t(32) = 2.061$, $p = 0.048$, $d = 0.027$].

A complementary analysis of participants' response times revealed a facilitatory effect of valid vs. neutral cues during 100% blocks [M = 0.997 vs. 1.220 sec, $t(32) = 21.83$, $p = 8.66e-21$, $d = 0.804$]. A repeated measures ANOVA applied to recall errors during 75% blocks revealed a significant effect of cue type (i.e., valid, invalid, neutral) [$F(2,66) = 85.433$, $p < 1e-31$; $\eta^2 = 0.728$; Figure 1D]. Analyses of cue effects on response times revealed a greater benefit from valid cues during 100% vs. 75% blocks significantly larger effect of valid cues during 100% vs. 75% blocks [M = 224 vs. 148 ms, respectively; $t(32) = 6.383$, $p = 3.597e-7$, $d = 1.093$]. Invalid cues during 75% blocks did not incur a response time cost compared to neutral trials (M = 20.5 ms; $t(32) = 1.702$, $p = 0.098$, $d = 0.084$). Thus, valid retrocues improved participants' recall errors and

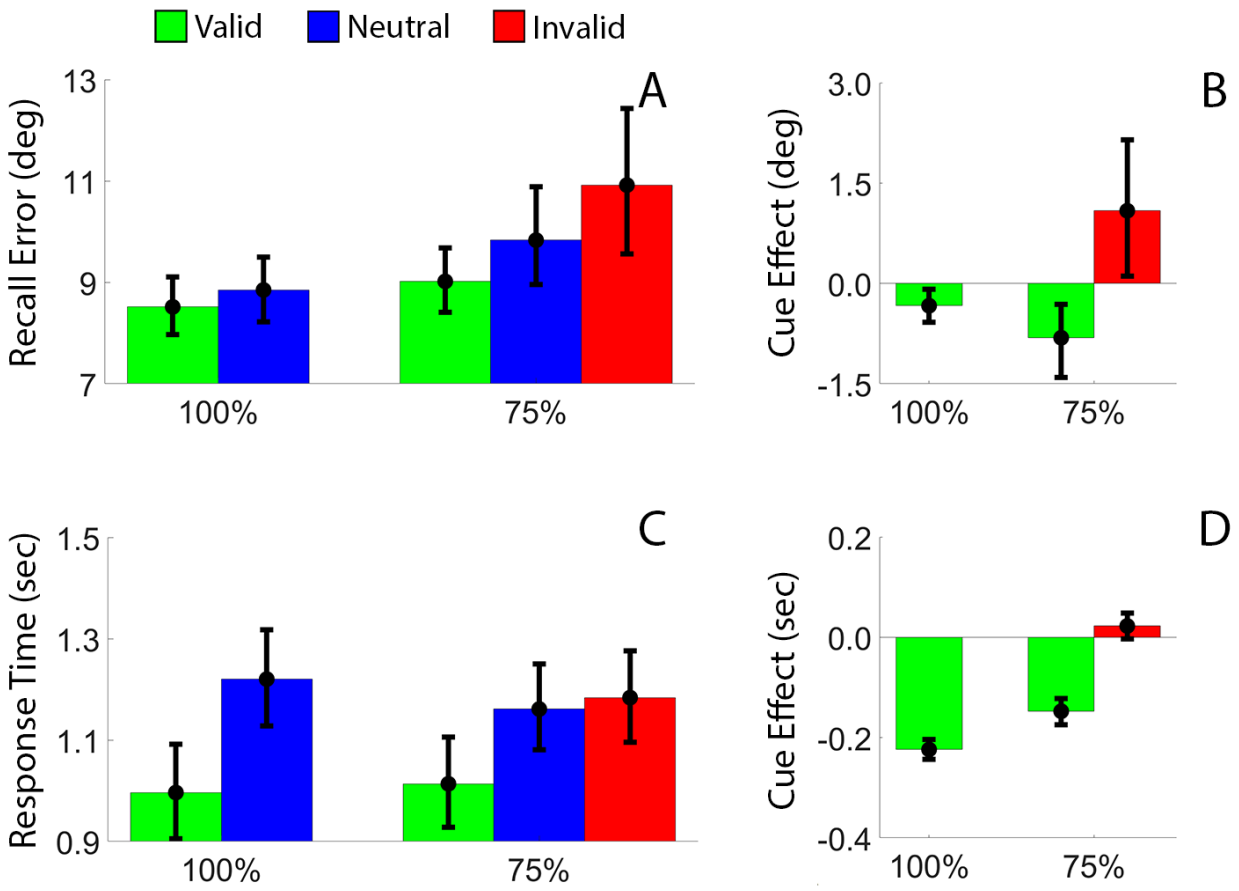

Figure 6. Behavioral Performance in Experiment 2. (A) Average absolute recall error as a function of cue type (valid, neutral, invalid) and cue reliability (100%, 75%). (B) Cue effects, defined as the difference between valid/invalid and neutral trials. (C) Average response times and (D) cue effects. Error bars depict the 95% confidence interval of the mean.

response times, and the magnitude of the response time benefit scaled with cue reliability (i.e., 100% vs. 75%).

EEG Decoding Performance. EEG data from Experiment 2 were analyzed in an identical way to those from Experiment 1. As in Experiment 1, we observed no effect of cue reliability on peak decoding accuracy or latency during neutral trials (Figure 7). Analyses of data from informative cue trials (Figure 8) were largely consistent with the findings of Experiment 1. During 100% blocks, decoding performance for the cue

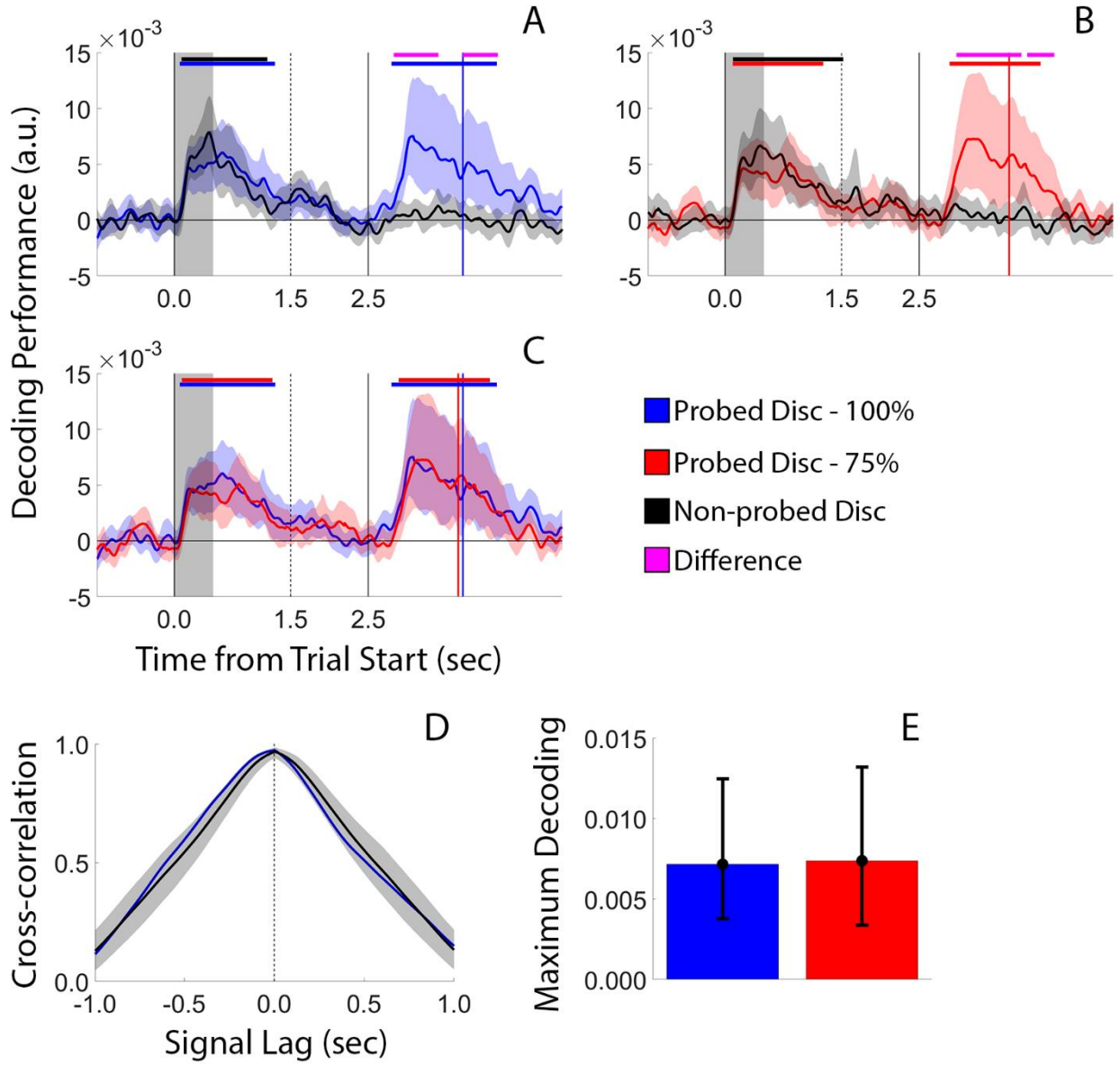

Figure 7. Decoding Performance During Neutral Trials of Experiment 2. Vertical lines at times 0.0, 1.5, and 2.5 sec mark the onset of sample, retrocue, and probe displays, respectively. All other conventions are identical to Figure 4.

matching and non-matching discs increased rapidly following onset of the encoding display but returned to chance levels by the onset of the retrocue display 1.5 seconds later (Figure 5A). Following the appearance of the retrocue, decoding performance for the cue-matching disc “recovered” to above-chance levels, replicating earlier findings showing cue-driven recovery in the quality of location-specific mnemonic

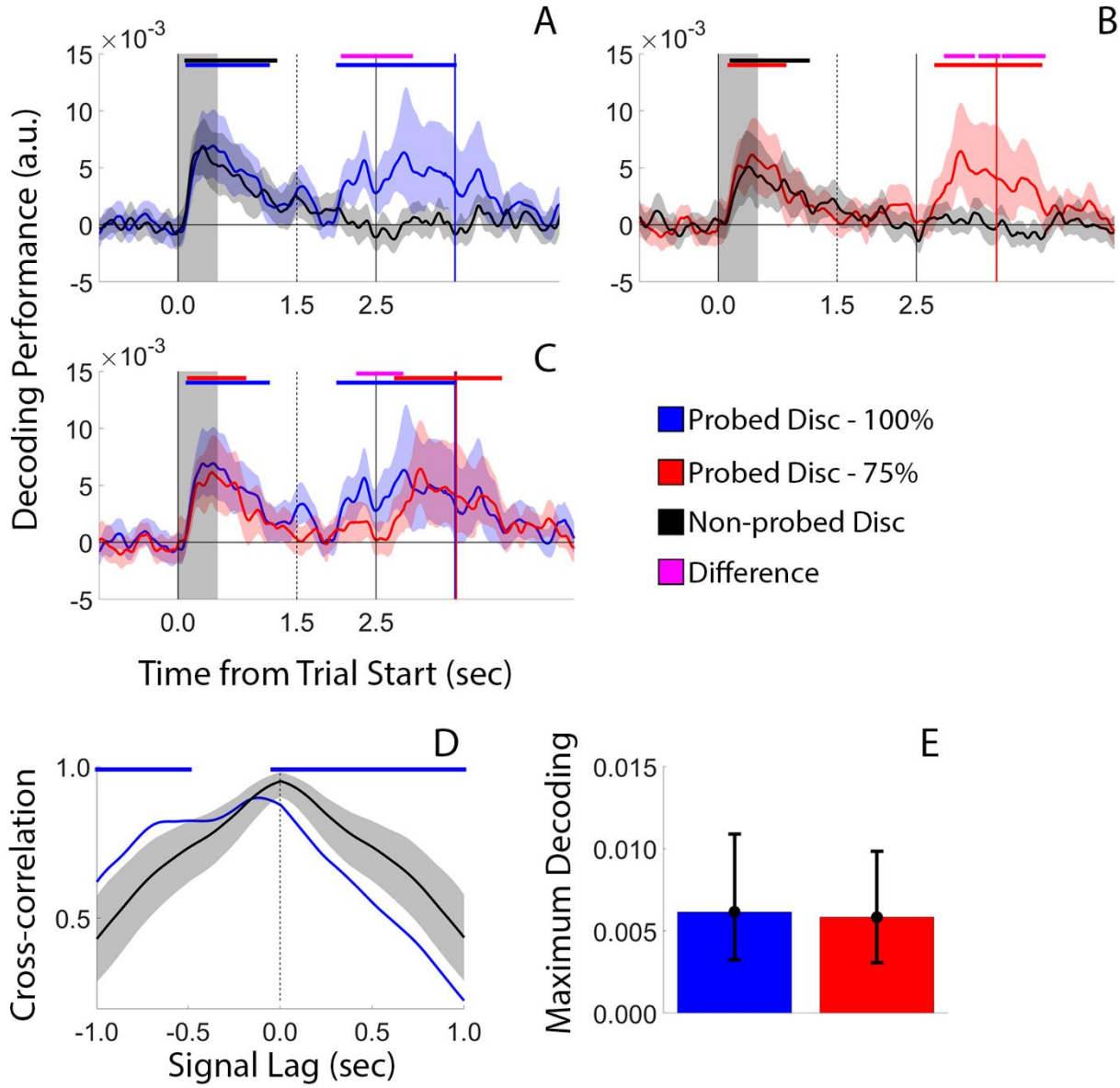

Figure 8. Decoding Performance During Informative Trials of Experiment 2. Vertical lines at times 0.0, 1.5, and 2.5 sec mark the onset of sample, retrocue, and probe displays, respectively. All other conventions are identical to Figure 5.

representations (Sprague et al., 2016; Ester et al., 2018). Decoding performance for the cue-matching item remained at above-chance levels through the probe display and until shortly after participants' responses. Conversely, decoding performance for the cue-nonmatching item remained at chance levels throughout the retrocue and probe displays.

An analysis of decoding performance during 75% blocks revealed a different pattern of findings (Figure 8B). Here, cue-matching and cue-nonmatching decoding performance remained indistinguishable from chance during the retrocue display, with cue/probe-matching decoding performance reaching above-chance levels only after the appearance of the probe display (as in Experiment 1, since invalid cues had no effect on participants' recall errors or response times, we pooled data from valid and invalid trials to create the data shown in Figure 8B). Time courses of cue- and probe-matching decoding performance during 100% and 75% blocks are shown in Figure 8C. Comparisons of peak decoding latency (Figure 8D) revealed that maximum decoding performance was reached significantly earlier during 100% blocks relative to 75% blocks, although average peak decoding performance did not differ across these conditions. Thus, the findings of Experiment 2 are qualitatively identical to Experiment 1: changes in the priority of location-specific WM representations influenced the latency but not the magnitude of peak decoding performance.

Alternative Explanations and Control Analyses. The data reported here suggest that access to cue-matching information is delayed when cue reliability is reduced (e.g., Figures 5D and 8D). An alternative explanation for this pattern of results holds that participants were simply more cautious or took more time to respond when cue reliability was fixed at 75% compared to when it was fixed at 100%. Analyses of participants response time do not support this possibility. Specifically, we reasoned that if participants were simply more cautious or slower in responding during 75% blocks, then their average response times during 75% valid trials should be significantly greater

than during 100% valid trials. In fact, response times following valid cues were identical during 100% and 75% blocks in both Experiment 1 (M = 1058 vs. 1068 ms, respectively; $t(33) = 0.579$, $p = 0.567$; green bars in Figure 3C) and Experiment 2 (M = 996 vs. 1013 ms, respectively; $t(32) = 0.768$, $p = 0.448$; green bars in Figure 6C). Thus, the findings reported here cannot be explained by general response caution or slowing during 75% vs. 100% blocks.

In both Experiments, cue-matching decoding performance during informative trials of 75% blocks fell to chance levels by the end of the memory period (Figure 5B and Figure 8B). This pattern is reminiscent of findings seen during neutral trials (Figures 4 and 7), raising the possibility that participants simply ignored the cues during 75% blocks. Once again, participants' memory performance argues against this claim: recall errors and response times were significantly lower during valid vs. neutral trials during 75% blocks (Figure 3C and 6C). Nevertheless, to investigate the possibility that participants simply ignored informative cues during 75% valid blocks we undertook analyses comparing the time-courses of cue-matching decoding performance during 75% valid trials and 75% neutral trials. If participants indeed ignored informative cues, then the time-courses of decoding performance should be identical during informative and neutral trials. Conversely, if participants used informative cues to prioritize relevant information stored in memory, then maximum decoding performance should be reached earlier during informative vs. neutral trials. We tested these possibilities using the same cross-correlation analyses used to quantify differences in maximum decoding latency during 100% and 75% blocks, and the results are summarized in Figure 9. Maximum decoding latency was reached significantly earlier during informative vs. neutral trials of

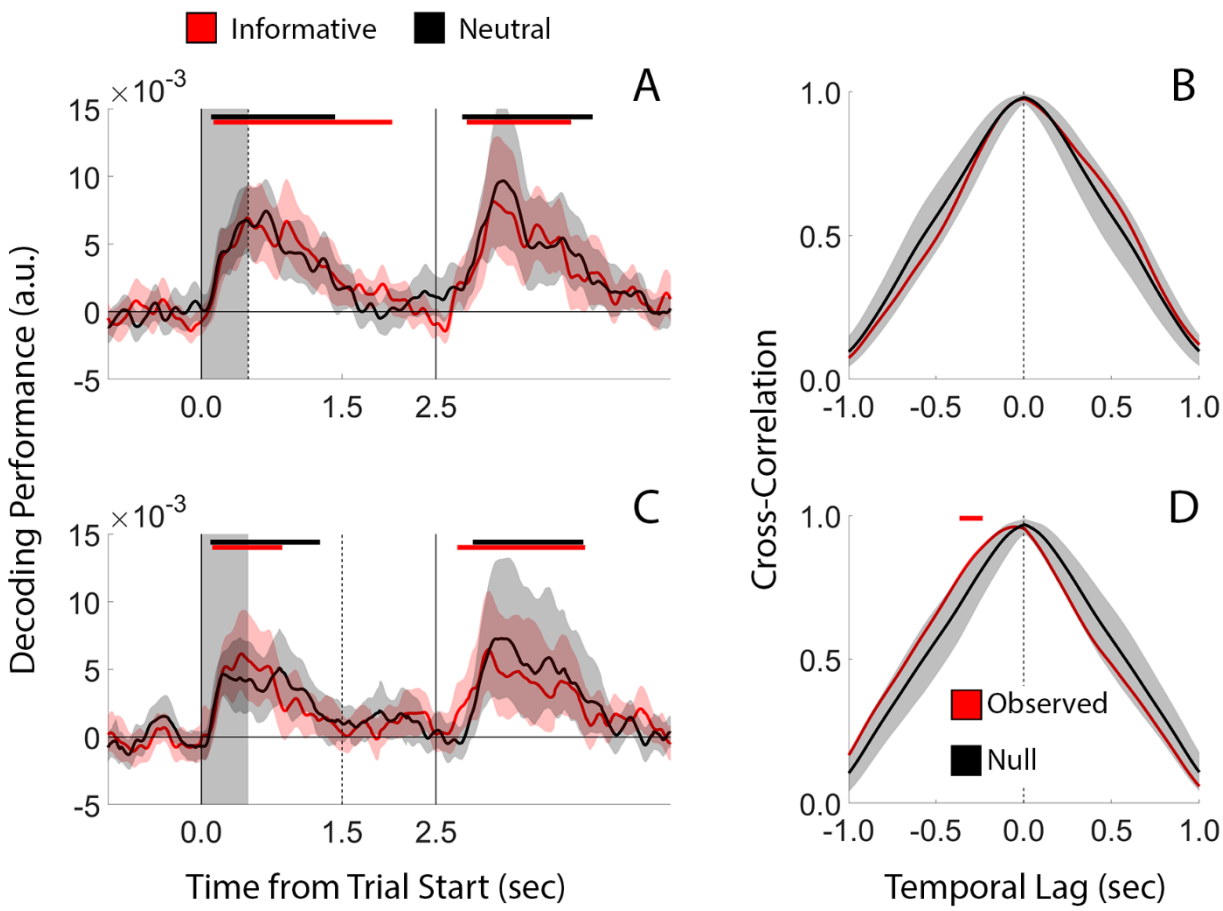

Figure 9. Comparisons of Above-Chance Decoding Latency on Informative and Neutral Trials during 75% blocks. (A) Overlay of above-chance decoding performance during informative and neutral cue trials of 75% blocks in Experiment 1. (B) Cross-correlation between probe-locked decoding performance during informative and neutral cue trials of 75% blocks in Experiment 1. Panels (C) and (D) are identical to panels (A) and (B) but use data from Experiment 2. Conventions are identical to those used in Figures 4-5 and 7-8.

75% blocks of Experiment 2 (Figure 9C and 9D), but not Experiment 1 (Figure 9A and 9B). These findings, coupled with analyses of participants' memory performance, provide converging evidence against the possibility that delays in achieving maximum decoding performance during 75% relative to 100% blocks (Figure 5D and 8D) were caused by participants simply ignoring informative cues during 75% blocks.

Discussion

Efficient behavioral selection requires rapid comparisons of sensory inputs with internal representations of motor affordances and goal states, and many of these comparisons take place in working memory (WM). Rapidly changing behavioral goals frequently require agents to assign (and re-assign) different levels of behavioral priority to items stored in WM. Prior studies utilizing retrospective cues suggest that humans can flexibly assign WM content with different levels of priority; for example, improvements in WM performance following an informative vs. uninformative retrospective cue scale positively with informative cue reliability (e.g., Berryhill et al., 2012; Shimi et al., 2014; Günseli et al., 2015; Günseli et al., 2019). Importantly, graded improvements in WM performance following a probabilistic retrocue could reflect (a) graded changes in the strength or quality of population-level, stimulus-specific neural patterns thought to mediate WM storage, (b) graded changes in how easily stimulus-specific neural patterns associated with high- (vs. low) priority items are accessed or "read-out" by downstream neural populations, or (c) some mixture or these options. Here, we leveraged the high temporal resolution of human EEG to adjudicate between these options. Specifically, we examined how graded changes in behavioral priority influenced our ability to decode a set of positions stored in WM. We reasoned that if priority-driven changes in memory performance are driven by the flexible allocation of attentional gain to items stored in memory, then population-level representations of high-priority items would be subject to larger gain modulations and thus easier to decode than low-priority items. Conversely, if priority-driven changes in memory performance are driven by changes in the accessibility of WM content, then it should be

possible to decode the identity of high-priority items at an earlier time than low-priority items. Our data support the latter view. Specifically, we found no evidence suggesting that changes in behavioral priority – as manipulated via post- or retrocue reliability – influenced asymptotic decoding performance for cue-matching memoranda. Instead, the latency of asymptotic decoding performance was significantly earlier for high-priority vs. lower-priority items (e.g., Figures 5 and 8). Thus, we conclude that changes in behavioral priority influence the accessibility but not the quality of stimulus-specific patterns of neural activity representing memoranda.

We have framed our discussion in terms of how changes in behavioral priority affect items that are likely to be probed for report, but some consideration should also be given to the effects of priority on items that are unlikely to be probed for report. For example, at least one previous study has reported that changes in behavioral priority affect the likelihood that cue-nonmatching items are attended and/or stored in WM (Günseli et al., 2019). The authors of this study used lateralized measures of covert spatial attention (alpha-band suppression; Klimesch, 2012) and WM storage (contralateral delay activity; Vogel & Machizawa, 2004) to show that cue-nonmatching items were less likely to be attended or retained in WM when cue reliability was low vs. high. Conversely, we found little evidence to suggest that changes in behavioral priority influenced neural patterns associated with memory for cue-nonmatching positions. However, a major difference between the current study and prior work that we included 100% reliable cue blocks, which allowed participants to simply drop or forget cue-nonmatching information if they chose to do so. Thus, it is difficult to directly compare cue-nonmatching decoding performance across different levels of cue reliability (i.e.,

100% vs. 75%). Additional traction on this issue could possibly be gained by comparing decoding performance for non-cued items during 75% blocks with decoding performance for both remembered items during neutral trials (akin to a 50% valid condition). However, we failed to observe above-chance decoding for the cue-nonmatching position during the delay period of 75% blocks or above-chance decoding for either remembered position during the delay period of neutral trials (e.g., Figures 4, 5B; 7, and 8B). Thus, additional research using a more fine-grained manipulation of cue reliability (e.g., 60%, 75%, 90%) will be needed to elucidate how changes in behavioral priority influence neural representations of cue-nonmatching items.

The current study utilized a mixture of postcues (Experiment 1) and retrocues (Experiment 2). Following earlier work (e.g., Souza & Oberauer, 2016) we use the term postcue to refer to any event informing which of a set of remembered item(s) will be tested that occurs immediately after encoding (including instances where an agent can apply this information to stimulus representations in sensory memory) and retrocue to refer to any informative event occurring after memory consolidation is complete. Prior evidence suggests that post- and retrocues engage separate but complementary mechanisms that promote storage of high-quality neural representations of memoranda. In an earlier study (Ester et al., 2018) we used an inverted encoding model to reconstruct location-specific representations of items stored in WM while presenting participants with neutral or perfectly informative postcues and retrocues. During neutral trials, the quality of reconstructed location-specific representations gradually decreased during WM storage. A perfectly valid postcue presented immediately after encoding eliminated this decrease, while a perfectly valid retrocue presented midway through

storage partially reversed it. Data from 100% blocks in this study replicate these findings (e.g., Figure 5A; Figure 8A) while also establishing that assigning lower priority to items stored in memory reduces these effects (Figure 5B; Figure 8B).

In the absence of a perfectly informative postcue (i.e., Figure 5A), our ability to decode the location of the cue-matching or nonmatching disc fell to chance levels during the delay period (e.g., Figures 4A-C; Figure 5B; Figure 7; Figure 8B). This does not imply a loss of memory: participants still performed quite well during neutral trials despite no evidence for above-chance decoding during the delay period (Figures 3-4 and 6-7). One possibility is that the amount of location-specific information carried by induced alpha patterns that were used for decoding in this study gradually decreases over time. This, however, is difficult to reconcile with observations of robust above-chance location decoding during recall, as the probe display contained no additional information about the location of the to-be-reported disc (i.e., the color of the fixation point instructed participants which disc to recall, but gave no additional information about its position at the beginning of the trial). To account for this pattern, we speculate that during WM storage position-specific memory representations are gradually consolidated into a new format not indexed by alpha-band activity storage (for example, in an “activity-silent” synaptic network or in long-term memory; Rose et al., 2016; Sprague et al., 2016; Wolff et al., 2017; Masse, 2019; Barbosa et al., 2021; Beukers et al., 2021), and later retrieved from this format during memory recall.

WM can be conceptualized as a temporal bridge between fleeting sensory phenomena and possible actions. Recent theoretical conceptualizations of WM have begun to emphasize the action-oriented nature of this system (e.g., Olivers &

Roelfsema, 2020; Heuer et al., 2020; van Ede & Nobre, 2022), and recent empirical findings suggest that behavioral (Ohl & Rolfs, 2020), circuit-level (Pho et al., 2018; Tang et al., 2020; Panichello & Buschman, 2021), and systems-level (Chatham et al., 2014; van Ede et al., 2019; Boettcher et al., 2021; Galero-Salas et al., 2021; Rac-Lubashevsky & Frank, 2021; Henderson et al., 2022) mechanisms of WM storage and action planning are tightly interwoven. In dynamic contexts where the future can take on several possibilities, the (potential) behavioral relevance of information stored in WM is often unknown. Thus, the likelihood that that any one piece of information stored in WM will become behaviorally relevant is best understood as a matter of probability rather than a certainty. From this perspective, a central purpose of WM may be to prepare for multiple potential futures, while mechanisms of internal attention act to select and prioritize relevant WM content as our predictions change or our uncertainties are reduced. The findings reported here are consistent with this view and further suggest that human observers can assign prospectively task-relevant representations different levels of priority that influences how quickly they can be retrieved and acted upon.

References

Barbosa J, Lozano-Soldevilla D, Compte A. (2021) Pinging the brain with visual impulses reveals electrically active, not activity-silent, working memories. PLOS Biology 19(10):e3001436

Bays PM, Taylor P. (2017) A neural model of retrospective attention in visual working memory. Cognitive Psychology, 100:43-52

Berryhill ME, Richmond LL, Shay CS, Olson IR (2012) Shifting attention among working memory representations: Testing cue type, awareness, and strategic control. Quarterly Journal of Experimental Psychology 64:426-438

Beukers AO, Buschman TJ, Cohen JD, Norman KA. (2021) Is activity silent working memory simply episodic memory? Trends in Cognitive Sciences 25:284-293

Boettcher, S.E.P., Gresch, D., Nobre, A.C., & van Ede, F. (2021). Output planning at the input stage in visual working memory. *Science Advances, 7*(13), eabe8212.

Chatham, C.H., Frank, M.J., & Badre, D. (2014). Corticostriatal output gating during selection from working memory. *Neuron, 81*(4), 930-942.

Chatham CH, Badre D (2015). Multiple gates on working memory. Current Opinion in Behavioral Sciences 1:23-31.

Cisek P (2019) Resynthesizing behavior through phylogenetic refinement. Attention, Perception & Psychophysics, 81:2265-2287

Cisek P, Kalaska JF (2010) Neural mechanisms for interacting with a world full of action choices. Annual Review of Neuroscience 33:269-298

D'Esposito, M., & Postle, B.R. (2015). The cognitive neuroscience of working memory. *Annual Review of Psychology, 66*, 115-142.

Delorme A, Makeig S (2004) EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. Journal of Neuroscience Methods 134:9-21

Dube B, Lumsden A, Al-Aidroos N (2018) Probabilistic retro-cues do not determine state in visual working memory. Psychonomic Bulletin & Review 26:641-646

Ester EF, Nouri A, Rodriguez L. (2018) Retrospective cues mitigate information loss in human cortex during working memory storage. Journal of Neuroscience 38:8538-8548.

Ester EF, Nouri A. (2022) Competition between endogenous and exogenous attention delays the selection of task-relevant working memory content. bioRxiv

Galero-Salas, Y., Han, S., Sych, Y., Voigt, F.F., Laurenczy, B., Gilad, A., & Helmchen, F. (2021). Sensory and behavioral components of neocortical signal flow in discrimination tasks with short-term memory. *Neuron, 109*(1), 135-148.e6.

Griffin IC, Nobre AC (2003). Orienting attention to locations in internal representations. Journal of Cognitive Neuroscience 15:1176-1194

Günseli E, van Moorselaar D, Meeter M, Olivers CNL (2015) The reliability of retro-cues determines the fate of noncued visual working memory representations. Psychonomic Bulletin & Review, 22:1334-1341.

Günseli E, Fahrenfort JJ, van Moorselaar D, Daoultzis KC, Meeter M, Olivers CNL (2019) EEG dynamics reveal a dissociation between storage and selective attention within working memory. Scientific Reports 9:13499

Henderson, M.M., Rademaker, R.L., & Serences, J.T. (2022). Flexible utilization of spatial- and motor-based codes for the storage of visuo-spatial information. *eLife*, 11:e75688.

Heuer, A., Ohl, S., & Rolfs, M. (2020) Memory for action: A functional view of selection in working memory. *Visual Cognition, 28*(5-8), 388-400.

Kleiner M, Brainard D, Pelli D (2007) What's new in Psychtoolbox-3? Perception 36:1-16

Klimesch W. (2012) Alpha-band oscillations, attention, and controlled access to stored information. Trends in Cognitive Sciences, 16:606-617

Landman R, Spekreijse H, Lamme VA (2003) Large capacity storage of integrated objects before change blindness. Vision Research 43:149-164

Lewis-Peacock JA, Drysdale AT, Oberauer K, Postle BR (2012) Neural evidence for a distinction between short-term memory and the focus of attention. Journal of Cognitive Neuroscience, 24:61-79

Libby, A., & Buschman, T.J. (2021). Rotational dynamics reduce interference between sensory and memory representations*. Nature Neuroscience, 24*, 715-726.

Maris E, Oostenveld R (2007) Nonparametric statistical testing of EEG and MEG data. Journal of Neuroscience Methods 164:177-190.

Masse NY, Yang GR, Song HF, Wang X-J, Freedman DJ (2019) Circuit mechanisms for the maintenance and manipulation of information in working memory. Nature Neuroscience 22:1159-1167

Myers NE, Stokes MG, Nobre AC (2017) Prioritizing information during working memory: Beyond sustained internal attention. Trends in Cognitive Sciences, 21:449-461

Nouri A, Ester EF (2020). Recovery of information from latent memory stores decreases over time. Cognitive Neuroscience 11:101-110

Ohl, S., & Rolfs, M. (2020). Bold moves: Inevitable saccadic selection in visual short-term memory. *Journal of Vision, 20*(2), 11.

Olivers, C.N.L., & Roelfsema, P.R. (2020). Attention for action in visual working memory. *Cortex, 131*,179-194.

Panichello M, Buschman T (2021) Shared mechanisms underlie the control of working memory and attention. Nature 592:601-605

Pho, G.N., Goard, M.J., Woodson, J., Crawford, B., & Sur, M. (2018). Task-dependent representations of stimulus and choice in mouse parietal cortex. *Nature Communications, 9*, 1-16.

Rac-Lubashevsky, R., & Frank, M.J. (2021). Analogous computations in working memory input, output and motor gating: Electrophysiological and computational modeling evidence. PLoS Computational Biology 17(6), e1008971.

Rose NS, LaRocque JJ, Riggall AC, Gosseries O, Starrett MJ, Meyering EE, Postle BR (2016). Reactivation of latent working memories with transcranial magnetic stimulation. Science 354:1136-1139

Shimi A, Nobre AC, Astle DE, Scerif G. (2014) Orienting attention within visual short-term memory: Development and mechanisms. Child Development 85:578-592

Souza AS, Oberauer K (2016) In search of the focus of attention in working memory: 13 years of the retro-cue effect. Attention Perception & Psychophysics, 78:1839-1860

Souza AS, Rerko L, Oberauer K (2016) Getting more from visual working memory: Retro-cues enhance retrieval and protect from visual interference. Journal of Experimental Psychology: Human Perception & Performance

Sprague TC, Ester EF, Serences JT (2014) Reconstructions of information in visual spatial working memory degrade with memory load. Current Biology 24:2174-2180

Sprague TC, Ester EF, Serences JT (2016) Restoring latent visual working memory representations in human cortex. Neuron 91:694-707

Tang, C., Herikstad, R., Parthasarathy, A., Libedinsky, C., & Yen, S.C. (2020). Minimally dependent activity subspaces for working memory and motor preparation in the lateral prefrontal cortex. *eLife, 9,* e58154.

van Ede, F., Chekroud, S.R., Stokes, M., & Nobre, A.C. (2019). Concurrent visual and motor selection during visual working memory guided behavior. *Nature Neuroscience, 22*(3), 477-483.

van Ede F, Nobre AC (2022) Turning attention inside out: How working memory serves behavior. Annual Review of Psychology

Vogel EK, Machizawa MG (2004) Neural activity predicts individual differences in visual working memory capacity. Nature 428:748-751.

Wolff MJ, Jochim J, Akyurek EG, Stokes MG (2017) Dynamic hidden states underlying working-memory-guided behavior. Nature Neuroscience 20:864-871.