Neural reinstatement tracks object-based attentional selection in working memory

Abbreviated title: Object-based attentional selection in working memory

Frida A. B. Printzlau[1,2], Nicholas E. Myers[1,2,3], Sanjay G. Manohar[1,2,4], Mark G. Stokes[1,2]

1. Department of Experimental Psychology, University of Oxford
2. Wellcome Centre for Integrative Neuroimaging, University of Oxford
3. School of Psychology, University of Nottingham
4. Nuffield Department of Clinical Neuroscience, University of Oxford

Corresponding author: Frida Printzlau, fridaprint@gmail.com

Number of pages: 40
Number of figures: 5
Word count: Abstract (231), Significance statement (119), Introduction (644), Discussion (1350)

The authors declare no competing financial interests.

Acknowledgements
This research was funded by a Biotechnology and Biological Sciences Research Council grant (BB/M010732/1) and James S. McDonnell Foundation Scholar Award (220020405) to Mark G. Stokes and by the NIHR Oxford Health Biomedical Research Centre. Frida A.B. Printzlau was funded by a Biotechnology and Biosciences Research Council studentship (BB/M011224/1) and the Oxford Interdisciplinary Biosciences Doctoral Training Partnership. Nicholas E. Myers was funded by the Wellcome Trust (grant 201409Z/16/Z) and University College Oxford. Sanjay G. Manohar was funded by the MRC clinician scientist fellowship (MR/P00878X) and Leverhulme research grant (RPG-2018-310).

ABSTRACT

Working memory (WM) is the ability to hold information in mind in the short-term and use it flexibly for behaviour. Not all items are represented equally in WM. Attention can be allocated to select and privilege relevant WM content. It is unclear whether attention selects individual features or whole objects in WM. Here, we used behavioural measures, eye-tracking and electroencephalography (EEG) to test the hypothesis that attentional selection spreads between an object's features in WM. Twenty-six participants (male and female) completed a WM task that asked them to recall the angle of one of two oriented, coloured bars after a delay while EEG and eye-tracking data was collected. During the delay, an orthogonal 'incidental task' cued the colour of one item for a match/mismatch judgement. On congruent trials (50%), the cued item was probed during memory recall; on incongruent trials (50%), the other memory item was probed. As predicted, selecting the colour of an object in WM brought other features of the cued object into an attended state as revealed by EEG decoding, oscillatory α-power, gaze bias and improved subsequent orientation recall performance. Together, the results build a case for object-based attentional selection in WM. Analyses of neural processing at recall revealed that the selected object was automatically compared with the probe, whether it was the target for recall or not, providing a potential mechanism for non-predictive cueing benefits in WM.

SIGNIFICANCE STATEMENT

Everyday behaviour depends on working memory (WM)—the ability to hold information in mind for an upcoming task. Attention can be focused to prioritise the most relevant object in WM. Most objects consist of multiple features, such as shape and colour. It is unclear whether attention selects entire objects in WM or only enhances the currently relevant feature. Here, we show that when people focus on a single feature of an object in memory, such as its colour, other features of that object are shifted into the focus of attention and become prioritised for subsequent decision-making. The results show that well-known object-based attention mechanisms exist for internal attention and highlight objects as a key unit of selection in WM.

Working memory (WM) allows us to maintain and process information over the short-term (Baddeley, 2003). Selective attention plays a key role in top-down control of WM by biasing processing toward task-relevant information for storage and retrieval (Baddeley, 2012; Gazzaley and Nobre, 2012).

Early evidence for internal attention toward WM content came from studies showing that retro-cueing relevant WM content after encoding improves memory performance (Griffin and Nobre, 2003; Landman et al., 2003). Retro-cued content is represented more strongly in neural patterns as decoded from blood oxygen-level dependent and EEG activity (Lepsien and Nobre, 2007; Lewis-Peacock et al., 2012; LaRocque et al., 2013) and retro-cues engage spatial attention toward the memorised position of the cued object (Poch et al., 2014; van Ede et al., 2019).

There is substantial overlap between behavioural and neural mechanisms associated with attention to stimuli in the environment and internal attention toward content in WM (Chun and Johnson, 2011; Kiyonaga and Egner, 2013; Panichello and Buschman, 2021). In external visual attention, object-based theories propose that objects are the key unit of attentional selection (Duncan, 1984; Desimone and Duncan, 1995). When attention is directed toward one object-feature, neural processing is enhanced for both task-relevant and task-irrelevant features of that object, suggesting perceptual attention automatically spreads between an object's features (O'Craven et al., 1999; Ernst et al., 2013).

Whether objects are similarly the unit of attentional selection in WM remains a topic of debate. Neural activity preferentially codes for task-relevant features during the WM delay, suggesting feature-based attention may enhance task-relevant features and/or remove irrelevant features from memory (Woodman and Vogel, 2008; Serences et al., 2009; Bocincova and Johnson, 2019). This is further supported by recent behavioural studies showing that participants' memory judgements are improved if a retro-cue informs them which feature dimension will be probed after a delay (Niklaus et al., 2017; Hajonides et al., 2020). Moreover, one study reported retro-cue benefits when the colour of one object and the orientation of another were cued as relevant, suggesting attention may enhance individual object-features in WM (Niklaus et al., 2017).

Others have found that attentional enhancement spreads between an object's features in WM. Using an 'incidental task' to retro-cue a single feature of a multi-feature WM object in a non-predictive way, Zokaei et al. (2014a, 2014b) showed that subsequent memory performance improved for another feature of the same object. Such obligatory retrieval of associated object-features is predicted by theoretical models of attention in WM that rely on associative pattern completion by neural attractors (Lansner et al., 2013; Manohar et al., 2019), unlike models that activate features individually (Bays and Taylor, 2018).

Behavioural paradigms alone face a challenge when it comes to assessing the relative priority status of features in WM during the maintenance delay. In the present study, we combined behavioural methods, EEG and eye-tracking to investigate whether attentional selection automatically spreads between an item's features in WM. Participants performed an orientation recall WM task with an 'incidental' delay task serving as a non-predictive colour retro-cue. Importantly, the cue did not require retrieval of any other features and did not predict which item would be probed for orientation recall. On congruent trials (50%), the cued item was also probed for recall; on incongruent trials (50%), the uncued item was probed.

If attention spreads between an item's features in WM, selection of a single feature during the delay should enhance memory performance and neural processing of other features of the same object. As predicted, selecting the cued colour for the incidental task shifted other features of the cued object into an attended state. This was revealed by improved orientation recall, enhanced orientation decoding from EEG, lateralized α-power, and gaze bias relative to the cued item's position. Analyses of neural dynamics at recall indicated that the cued item was prioritized for comparison with the probe, providing a potential mechanism for non-predictive cueing benefits in WM.

Material and methods

The hypotheses and methods were pre-registered on the Open Science Framework after data collection had begun, but prior to data analyses (https://osf.io/pgdfj). The pre-processing pipeline was tested on a pilot participant not included in the analyses. The methods are as described in the pre-registration, unless otherwise specified in text and under 'deviation from pre-registered methods' section. Most notably, we planned to recruit a total of 35 participants, but due to Covid-19 restrictions only 32 participants completed the experiment.

Participants

A total of 32 participants completed the experiment (see *Deviations from pre-registered methods,* 1). One participant was excluded from analyses due to colour-blindness. A further five participants were excluded as they had less than 80% of usable epochs remaining after artefact rejection, in line with the pre-registered exclusion criteria, leaving a total of 26 participants for analyses (14 female, 12 male). Participants were aged between 18 and 34 (M=23.23, SD= 5.02). Twenty-two were right handed and four were left handed. All participants had normal or corrected-to-normal vision (including normal colour vision), and did not have any history of neurological or neuropsychiatric disorders. Participants provided informed consent prior to participating in the study. They were reimbursed for their time at a rate of £15/hour or with course credit. The study has been approved by the Oxford Central University Research Ethics Committee (R55073/RE008).

Procedure

Participants performed a computerised experimental task (Figure 1A). It consisted of a continuous report WM task, in which participants reported the orientation angle of one of two items in memory at the end of a delay. During the delay, participants performed a secondary "incidental task", where a non-predictive colour cue, presented centrally, required a speeded mismatch response: does the colour differ from the colours of the two memory items? At the end of the trial, a probe indicated which of the two items participants should report. This was either the same item as the one that was cued on the incidental task (congruent condition, 50% of trials) or the other item in memory (incongruent condition, 50% of trials).

Following an initial fixation period of 300 ms, participants were presented with a WM display consisting of two oriented, coloured bars. The WM items were presented on screen for 250 ms followed by a delay period of 750 ms. Then, for the incidental task, participants were shown a

central coloured ring as a cue for 200 ms. Participants indicated whether the colour of the cue matched the colour of either of the bars in memory. On mismatch trials (20%), participants pressed the space bar within 1000 ms of incidental cue onset, after which the rest of the trial was skipped. Participants received feedback for correct responses or incorrect omissions, in the form of a happy or sad smiley face for 200 ms. On match trials (80%), participants made no response and the trial continued (unless they accidentally pressed the space bar, in which case the trial was also discontinued). After another delay of 500 ms, the probe appeared on the screen. The colour of the probe indicated which of the two items participants should report. Participants used the arrow keys to turn the on-screen dial to match the orientation of the probed item and pressed the space bar to respond. There was no maximum time to respond, but once participants initiated a response the trial would time-out within 2500 ms. They received visual feedback of the correct orientation for 100 ms. The inter-trial-interval was uniformly jittered between 400-800 ms. At the end of each block, they received feedback of their incidental colour change-detection rate and on their mean absolute error on the orientation recall task.

Participants completed a total of 1400 trials. The 20% non-match incidental trials were excluded, leaving 1120 trials per participant for analyses. Participants practiced the task beforehand until they reached a criterion of >80% correct incidental task change-detection rate and <25 degrees mean absolute error relative to the target orientation.

Apparatus and stimuli

The task was programmed and stimuli presented in Matlab with Psychophysics Toolbox (Brainard, 1997; The MathWorks Inc., 2017). The task was presented on a 24-inch, 1920x1080 pixel monitor running at 100Hz.

Stimuli were presented on a grey background (RGB=128, 128, 128; see *Deviations from pre-registered methods,* 2). A fixation target was presented at the centre of the screen throughout the trial. It consisted of a filled black circle (0.5° diameter) with a white cross inside (width: 0.15°). To signal trial onset and to encourage fixation, a black dot (0.15° diameter) appeared inside the fixation target 300 ms before the onset of the memory array. The memory array consisted of two oriented, coloured bars inside circles (diameter: 6°, width: 0.3°). The colours were picked at random without replacement out of four possible colours on each trial: red (RGB=153, 0, 27), purple (RGB=122, 39, 150), green (RGB=16, 83, 0), and blue (RGB=0, 72, 136), chosen to be isoluminant in CieLab space. On every trial, the orientation of each

stimulus was randomly selected without replacement from a uniform distribution of 160 orientation angles between 1.1250 and 180 degrees. The two WM items were presented laterally 6° from the fixation target. For the incidental task, participants were presented with a coloured ring around the fixation target (diameter: 1°, width: 0.5°). After the memory delay, the probe was a tilted bar inside a circle (6°; same size as memory items) and matched the colour of the probed memory item. The orientation angle of the probe was selected at random on each trial between 1 and 180 degrees. The probe was presented 6° above or below the fixation target, with equal probability, pseudorandomised for the full experimental session. This type of probe display has been reported to reduce the risk of stimulus-specific eye-movements (Mostert et al., 2018). Feedback of the correct target orientation was presented as a white line on top of their reported orientation for comparison.

EEG data acquisition

Electroencephalography (EEG) was recorded at a sampling rate of 1kHz using Synamps amplifiers (Neuroscan, Inc., Albany, CA) and Curry Neuroimaging Suite (v7.0.12 X) software. Sixty-one electrodes were positioned according to the 10-10 system (American Clinical Neurophysiology Society, 1991). Eye movements were recorded from bipolar EOG derivations from electrodes above and below the right eye (vertical) and lateral to each eye (horizontal). A ground electrode was placed on the left elbow and the left mastoid was used as the active reference.

In addition, participants' eye movements were recorded using an eye tracking camera (EyeLink) tracking at least one eye at a rate of 1kHz. Eye-movements from six participants were accidentally recorded at a sampling rate of 500 Hz instead of 1000 Hz. These participants were not excluded as the eye-tracking data was down-sampled to 500 Hz for all participants, in line with the pre-registered pre-processing pipeline.

Pre-processing of eye-movement data

The EyeLink EDF file was converted to ASCII format and read into Matlab using the Fieldtrip toolbox (Oostenveld et al., 2011). The data were epoched around the onset of the WM array (-1 to 4 seconds). We took the mean of the right and the left eye (or used a single eye, where it was only possible to record from one eye; see *Deviations from pre-registered methods,* 3) and down-sampled the data to 500 Hz. Eye-blinks were detected (zeros in the eye-position data) and interpolated from -100 to +100 ms around the blink onset and offset using linear interpolation. Saccades (including microsaccades) were detected using a Matlab function (Engbert and

Mergenthaler, 2006) based on two-dimensional velocity vectors computed from eye-position data using moving averages (velocity type=2). We used a velocity threshold (λ) of 6 standard deviations of the velocity distribution, and minimum duration of 6 ms.

EEG Pre-processing

Pre-processing of the EEG data was performed in Matlab using the Fieldtrip toolbox (Oostenveld et al., 2011). The ERPs were re-referenced offline to the mean of the right and left mastoid and down-sampled to 500 Hz. A high-pass filter of 0.1 Hz and a low-pass filter of 40 Hz was applied to all channels. The data were epoched around the onset of the WM array (-1 to 4 seconds). We performed ICA to remove any heartbeat artefacts and/or artefacts due to eye-blinks. Artefact detection and rejection was performed separately for the time-window following WM array onset (0-1 s) and following cue onset (0-1.5 s). In addition, for an exploratory analysis not included in the pre-registration, we performed artefact rejection separately for a time-window following probe onset (0-1.5s). When an artefact was detected, we rejected the epoch containing the artefact from analyses, not the entire trial. We marked epochs with a high degree of variance using the Fieldtrip summary plot for visual artefact rejection and further marked epochs that exceed a cut-off threshold of 50 of the z-transformed value of the pre-processed data to detect any large artefacts due to muscle movements. We further excluded epochs with saccades >1° visual angle identified from the eye-tracking data. In case eye-tracking data was missing, we visually inspected EOGs to reject saccade epochs. Six participants had missing eye-tracking data in the range of 4-32 trials.

We excluded participants with fewer than 80% of epochs remaining following artefact rejection from further analyses. On average, we rejected 7.22% of trials from the encoding epoch, 10.20% of trials from the cue epoch and 2.13% of trials from the probe epoch.

Experimental design and statistical analyses

Behaviour

The behavioural measures of interest were mean absolute error from the target angle and median response initiation times (henceforth RT). If selecting the cued colour during the maintenance delay privileges associated but currently irrelevant item-features, subsequent orientation judgements should be faster and more precise on congruent than on incongruent trials. This would replicate previous behavioural results (Zokaei et al., 2014a, 2014b).

Alternatively, if it is possible to selectively attend the cued colour in WM without strengthening memory representations for associated features of the same object, there should be no difference in performance between congruent and incongruent trials. We performed paired samples t-tests (one-tailed) to test for a main effect of congruency for both measures. We predicted that mean absolute error and median RTs would be significantly lower for congruent relative to incongruent trials.

To investigate sources of error in memory judgements and how these were affected by cueing, we fit a swap model to the data (Bays et al., 2009). The model is an extension of a mixture model (Zhang & Luck, 2008) which aims to identify the frequency of recalling the probed item, erroneously recalling the other item ('swaps'), and random guessing. It can be described in the following way:

$$P(\theta') = (1 - \gamma - \beta)\phi_\sigma(\theta' - \theta) + \gamma\frac{1}{2\pi} + \beta\frac{1}{m}\sum_{i}^{m}\phi\,(\theta' - \theta_i^*)$$

Where θ denotes the target orientation in radians, θ' denotes the reported orientation, γ is the proportion of trials where the participant responds at random and ϕ_σ denotes the circular equivalent of the normal distribution (Von Mises distribution) with mean = 0 and standard deviation = σ. The swap model extension also includes the β parameter, which is the probability of responding with one of the non-target values: $\{\theta_1^*, \theta_2^*, \theta_m^*\}$ for m non-targets. As such, the model estimates the precision, target rate, non-target rate and guess rate of responses. As above, we performed paired samples t-tests to test for a main effect of congruency for each measure (one-tailed; congruent>incongruent for target rate and precision, and incongruent>congruent for guess rate and non-target rate; see *Deviations from pre-registered methods,* 4).

Orientation decoding from EEG voltages

If selection of the cued colour reactivates other features of the cued item even though these are irrelevant to the incidental task, we expected to be able to decode the orientation of the cued item from EEG activity patterns following cue onset and decoding quality to be greater for the cued than the uncued item's orientation. Alternatively, it may be possible to selectively access a single feature in memory, similar to selective coding for only relevant item features observed during encoding (Serences et al., 2009; Bocincova and Johnson, 2019). If so, patterns of

activity in the EEG channels should contain no information about the cued item's orientation and/or contain equal information about the orientations of both items following the cue.

To test whether the patterns of activity in the EEG channels contained information about the orientation angles of the items in memory, we used Mahalanobis distance to calculate the trial-wise distances between the multidimensional activity at each time point for each of the 160 possible orientation angles in memory, grouped into 16 angle bins distributed around a 180 degree angle space. The left and right memory items were decoded separately and independently for each participant. We performed the analyses on the 17 posterior channels (P7, P5, P3, P1, Pz, P2, P4, P6, P8, PO7, PO3, POz, PO4, PO8, O1, Oz and O2) as in (Wolff et al., 2017).

The decoding procedure followed an 8-fold cross-validation approach to calculate the decoding accuracy of the orientation of interest for each trial. The activity pattern of the trials of the testing fold at a particular time-point were compared to the trials of the 7 training folds. The trials of the training folds were averaged into 16 orientation bins relative to the test trial orientation, each bin containing trials with orientations within a range of 11.25°. As trials were rejected due to artefacts, we ensured equal trial numbers in each bin by random subsampling. We computed the pairwise Mahalanobis distances between the test trials and each of the orientation bins using the covariance matrix estimated from trials in the 7 training folds, using a shrinkage estimator (Ledoit and Wolf, 2004). To obtain a visual representation of a tuning curve, the 16 distances were ordered as a function of orientation difference, mean-centered, and sign-reversed, so higher values represented greater similarity between the test trial and the training set. This was repeated for all train and test fold combinations. To obtain reliable estimates, the above procedure was repeated 100 times (with random folds each time), separately for eight orientation spaces used for binning the training trials with respect to the test trial (bin centres: 1.41°-170.16°, 2.81°-171.56°, 4.22°-172.97°, 5.63°-174.38°, 7.03°-175.78°, 8.44°-177.19°, 9.84°-178.59° and 11.25°-180°, each in steps of 11.25°). The resulting 800 samples (100 repetitions×8 orientation spaces) were averaged for each of the 16 Mahalanobis distance values at each time point.

Finally, to obtain a summary measure of decoding accuracy, we computed the cosine-weighted means of the tuning curves (Sprague et al., 2016; Wolff et al., 2017). Higher values reflect greater orientation tuning and thereby greater decoding quality for that trial and chance-level

is zero. We took the mean over trials, resulting in a single decoding value for each time-point for each participant.

To test for significant orientation decoding of the two items, we took the mean decoding accuracy within the time-window following array onset (0–1 s; WM array epoch) and following cue onset (0–1 s; cue epoch). As a positive control we tested if the orientation of both items could be decoded during the WM array epoch using one-sample t-tests (one-tailed>0) and tested for any differences in decoding strength between items using a paired sample t-test (two-tailed). We expected to be able to decode the orientation of both items during the WM array epoch and decoding quality to be similar for both items.

The primary analysis of interest tested whether the orientation of the cued item could be decoded during the cue epoch. Again, we used one sample t-tests, to test for significant decoding of each item (one-tailed, >0) and a paired sample t-test (one-tailed, cued>uncued) to test for any difference in decoding for the two items. We expected orientation decoding to be significant for the cued item and decoding quality to be significantly greater for the cued than uncued item.

For completeness, we also tested for significant decoding of each memory item and for any difference between them at each time-point (one-sample t-test, one-tailed, >0, paired samples t-test, two-tailed) using a cluster-based permutation test (10,000 permutations) to correct for multiple comparisons over time. We used a cluster-forming and cluster significance threshold of $p<.05$. Time-resolved decoding values were smoothed with a Gaussian kernel (SD=10 ms) for visualisation and statistical tests.

Spatiotemporal decoding of orientation from EEG

Decoding accuracy may be improved by taking advantage of the dynamic nature of ERPs (Grootswagers et al., 2016; Wolff et al., 2020). This approach pools the relative voltages over time and sensors, thus leveraging information encoded in temporal as well as spatial patterns. Building on the pre-registered methods, we complement the standard decoding with this spatiotemporal decoding approach focusing on a 400 ms time window from 100-500 ms following onset of the event of interest (WM array onset and cue onset). First, the data was down-sampled to 50 Hz by taking the average every 10 ms. We then zero-centered the voltage fluctuations within the time-window by subtracting the mean voltage over the full time-window from the voltage at each time point. This relative baseline is comparable to pre-

stimulus baseline correction, common in ERP analyses, but further removes stable neural activity within the time-window and is optimised for decoding stimulus-evoked, dynamic activity. The resulting 40 values for each posterior channel were concatenated and used as the input to the multivariate decoder. The decoding procedure followed the same steps described above, but resulted in a single decoding value for each time-window of interest. As above, the procedure was repeated 100 times (with random folds each time), separately for eight orientation spaces and averaged across the resulting 800 samples for each of the 16 Mahalanobis distance values at each time-point.

As above, we used one sample t-tests, to test for significant decoding of each item (one-tailed, >0). During the WM array epoch, we expected decoding strength to be significant for both items. During the cue epoch, we expected orientation decoding to be significant for the cued item and decoding strength to be greater for the cued than the uncued item (one-tailed; cued>uncued).

Time frequency decomposition and alpha lateralization

Using alpha lateralization and gaze bias as indices of spatial attention in WM, we tested if the centrally presented, non-predictive colour cue engaged spatial attention toward the original position of the cued item, despite location being a fully redundant feature in this task.

If so, we expected contralateral minus ipsilateral alpha power to be reduced relative to the original location of the cued item. We computed the alpha-band power lateralization index relative to the cued item's location in the time window after incidental cue onset (0.2–0.8 s; to avoid effects related to probe onset). Spectral decomposition of the data was performed in Matlab with the FieldTrip toolbox (Oostenveld et al., 2011). We applied Hanning tapers with a time window width of five cycles per frequency, with frequencies of interest between 8 Hz and 14 Hz in steps of 1 Hz. We log-transformed the power at each frequency $10 \times log_{10}$ and calculated the mean power over contralateral and ipsilateral posterior channels (P8/7, P6/5, P4/3, PO8/7, PO4/3, O2/1) relative to the cued item. To get a lateralization index, we took the difference between contralateral and ipsilateral alpha power averaged over the specified time-window (0.2-0.8s) and tested for significant lateralization using a one-sample t-test (one-tailed, <0; see *Deviations from pre-registered methods,* 5). In an exploratory analysis, we also tested for significant alpha lateralization at each time-point across the time-window from 0-1 s (one-sample t-test, two-tailed) using a cluster-based permutation test (10,000 permutations) to

correct for multiple comparisons over time. We used a cluster-forming and cluster significance threshold of $p<.05$.

Gaze bias analysis

Human gaze may be biased toward attended spatial locations in WM (van Ede et al., 2019; Van Ede et al., 2020). Here, we similarly used gaze bias as an index of spatial attention towards the original position of the cued item. If spatial attention is directed toward the original position of the cued item, we expected horizontal gaze position to be biased toward the left (i.e., negative values) for left cue trials and toward the right (i.e., positive values) for right cue trials. First, we centred the gaze position data to the mean position during fixation (300–100 ms before stimulus onset). We then normalized the data according to the eccentricity of the memory items, so $+/-100\%$ corresponds to the eyes being focused on the original position of the memory items (i.e., 6° visual angle) and 0% corresponds to fixation. We plotted the gaze position following incidental cue onset separately for left and right cue trials. Heat maps of gaze density were created by computing the 2D histograms (bin size=.012 dva) of eye-position from 0.2-0.8 s from cue onset separately for the two possible locations of the cued memory item. We converted histogram counts to density by dividing by the trial number for each location condition before subtracting out density common to both locations. Finally, we took the mean across participants (see *Deviations from pre-registered methods,* 6*).*

To obtain an aggregate measure of 'towardness', we took the mean horizontal gaze position for right cue trials minus the mean gaze position for left cue trials, divided by two, averaged across the time-window following incidental cue onset (0.2–0.8) (van Ede et al., 2019). We tested for significant gaze bias toward the cued item using a one-sample t-test (one-tailed, >0). In an exploratory analysis, we also tested for significant towardness at each time-point across the time-window from 0-1 s (one-sample t-test, two-tailed) using a cluster-based permutation test (10,000 permutations) to correct for multiple comparisons over time. We used a cluster-forming and cluster significance threshold of $p<.05$. Time-resolved towardness values were smoothed with a Gaussian kernel (SD=10 ms) for visualisation and statistical tests.

Deviations from pre-registered methods

1. We had planned to recruit a total of 35 participants, but due to Covid-19 restrictions only 32 participants completed the experiment and 26 participants were included in the analyses following exclusion as per the pre-registered criteria.

2. The pre-registration mistakenly stated stimuli would be presented on a grey background with RGB values: 0.5, 0.5, 0.5, when the correct RGB values are: 128, 128, 128.
3. The pre-registration stated we would take the mean of the right and the left eye position. Due to some participants wearing glasses, it was only possible to record from one eye. We did not exclude these participants, but used the position data from the single recorded eye in these cases.
4. When testing for an effect of congruency on the parameters of the mixture model, the pre-registration mistakenly stated 'two-tailed' paired samples t-tests would be performed. As we specifically predicted performance to be better for congruent than incongruent trials, we performed one-tailed paired samples t-tests to test for a main effect of congruency for each measure (congruent>incongruent for target rate and precision, and incongruent>congruent for guess rate and non-target rate).
5. When testing for significant alpha lateralization following cue onset, the pre-registration specified the wrong direction of the one-tailed test ('>0' should be '<0'). We were specifically interested in contralateral alpha suppression, so we expected contra–ipsilateral alpha power to be negative.
6. The precise methods for plotting the eye-position heat maps were not included in the pre-registration, but are included in the methods for clarity.

Exploratory analyses

Relationship between cue-locked orientation decoding and behaviour (error/RT)

Representational quality of the cued item's orientation may signal the extent to which the cued item enters a privileged state in WM. If so, greater orientation decoding quality may be associated with a larger behavioural congruency benefit. We computed the Spearman correlation coefficient between the participants' mean orientation decoding quality across the cue epoch (0-1s after cue onset) and the congruency benefit on mean absolute error and RT (congruent – incongruent). We expected participants with greater orientation decoding quality to show lower error/RT on congruent relative to incongruent trials (rs<0).

To test whether cued orientation decoding quality was specifically related to orientation recall performance on congruent or incongruent trials, we additionally computed the Spearman

correlation coefficient between participants' orientation decoding quality and error/RT separately for congruent and incongruent trials. A negative correlation on congruent trials (rs<0) would suggest that greater decoding quality for the cued item's orientation facilitates performance (i.e., reduces error) when the same item is probed for recall. A positive correlation on incongruent trials (rs>0) would suggest that greater decoding quality for the cued item's orientation impairs performance when the other WM item is probed for recall. However, we expect overall task performance to be a less sensitive measure than the congruency benefit, as a large proportion of variability is likely explained by task-extraneous factors, such as alertness or age.

In addition to the subject-wise correlations, we tested for a trial-wise relationship between orientation decoding quality and behavioural performance. Two general linear models were fitted for each participant with trial-wise orientation decoding quality averaged over the time-window of interest (0-1 s from cue onset) as the predictor variable. The outcome variable was log-transformed RTs in one model and absolute error in the other. The design matrices included a constant term. We fit the models separately for congruent and incongruent trials and obtained β weights for each participant. We performed one-sample t-tests on the β values. We expected greater decoding quality following the cue to predict lower RT/error on congruent trials (one-tailed, <0) and higher RT on incongruent trials (one-tailed, <0).

To test whether potential relationships to behaviour were specific to particular time-points in the post-cue time-window, we fit the regression models over time to obtain β weights for each time point during the delay following cue onset and performed cluster-corrected t-tests across a time-window from 0-1 s following cue onset (two-tailed, cluster-forming threshold=.05, 10,000 permutations). Time-resolved β values were smoothed with a Gaussian kernel (SD=10 ms) for visualisation and statistical tests.

Relationship between cue-locked alpha lateralization/gaze bias and behaviour (error/RT)

We tested for subject-wise and trial-wise relationship between alpha lateralization/gaze bias and behavioural orientation recall performance (error/RT) using the same methods as described for orientation decoding above.

For the subject-wise correlations, we expected participants with greater contralateral alpha suppression to show a larger congruency benefit, i.e., lower error/RT on congruent relative to

incongruent trials (rs>0). We expected the reverse pattern for gaze bias: participants with greater gaze bias should show a larger congruency benefit (rs<0).

We performed one-sample t-tests on the trial-wise β values. We expected contralateral alpha suppression to predict lower RT/error on congruent trials (one-tailed, >0) and higher RT/error on incongruent trials (one-tailed, >0). We expected greater gaze bias following the cue to predict lower RT/error on congruent trials (one-tailed, <0) and higher RT on incongruent trials (one-tailed, <0).

Probe processing

At orientation recall, participants compared the orientation of the randomly oriented probe stimulus to the probed item in memory in order to turn the dial to the appropriate orientation angle (target). The neural representation of the absolute distance between the random probe orientation and the memory item may signal how well the memory item is compared to the probe at recall. In this analysis, we aimed to test whether computation of this comparison signal was modulated by whether the probed item in memory was previously cued (congruent) or not (incongruent).

Decoding target-probe distance prior to response onset

We performed the Mahalanobis distance decoding procedure specified above on the absolute angular distance between the target item in memory and the randomly oriented probe presented at orientation recall (target-probe distance). Instead of sorting the trials into 16 bins from 0-180°, the trials were averaged into 8 angular distance bins from 0-90°, each bin containing trials with orientations within a range of 11.25°. Otherwise the decoding procedure followed the same 8-fold cross-validation approach to calculate the decoding accuracy of the absolute angular distance of interest for each trial, repeated 100 times with random folds each time. Finally, as the absolute angular distance had a bounded uniform distribution rather than circular distribution, we computed the linear slope of the tuning curve instead of the cosine-weighted means to obtain a summary measure of decoding quality (Muhle-Karbe et al., 2021).

Training and testing was done separately for congruent and incongruent trials. The analysis was performed on the time-window from –0.5-0 s before the response onset (response-locked). Since reaction times were mostly longer than 0.5 s (72.2% of trials), the response-locked epoch mostly contained time points after the onset of the probe (Figure 5G). One-sample t-tests were

performed to test for significant target-probe distance decoding for congruent and incongruent trials within in time-window of interest (–0.5-0s; one-tailed, >0) and a paired-samples t-test was performed to test for a significant difference between congruent and incongruent trials (two-tailed). Additionally, cluster-corrected t-tests were performed across the full time-window prior to response onset (–1-0 s; two-tailed, cluster-forming threshold=.05, 10,000 permutations). Decoding time courses were smoothed with a Gaussian kernel (SD=15 ms) for visualisation and statistical tests.

Decoding target-probe distance and nontarget-probe distance following probe onset

To preview the results, we found significant decoding of the target-probe distance prior to response onset and decoding quality was greater on congruent than incongruent trials. We next asked whether, on incongruent trials, the probe might initially be compared to the cued item, even though it is the non-target. To capture the probe-evoked neural response, the analysis was time-locked to the probe-onset (0-0.5s) rather than to response onset as above. Using the same decoding procedure described above, we decoded the target-probe distance and the nontarget-probe distance. Note that on congruent trials, the cued item is the same as the target, whereas on incongruent trials, the cued item is the nontarget. Thus, decoding strength of the nontarget-probe distance on incongruent trials may indicate the extent to which the cued item is incorrectly used as a template for comparison with the probe.

One-sample t-tests were performed to test for significant decoding of the target-probe and nontarget-probe distance for congruent and incongruent trials and for significant differences between them within the time-window of interest (0-0.5s; one-tailed, >0). Cluster-corrected t-tests were performed across the full time-window following probe onset (0-0.5s; two-tailed, cluster-forming threshold=.05, 10,000 permutations). Time-resolved decoding values were smoothed with a Gaussian kernel (SD=15 ms) for visualisation and statistical tests.

Significant nontarget-probe distance coding could be a generic comparison that is unrelated to performing the task and preparing a response. To test the extent to which the neural pattern coding for the nontarget-probe distance resembled the neural pattern coding for the target-probe distance on congruent trials, we repeated the same analysis, but instead of training and testing the decoder within each condition, the training data were congruent trials labelled with target-probe distance values, while the test data were either congruent or incongruent trials labelled with non-target probe distance values. By training on the same data (target-probe distance), we have a better chance of comparing the same neural mechanism. On

congruent test trials, 8-fold cross validation was performed as above. As the training and test data were independent on incongruent test trials, we did not perform 8-fold cross-validation. Significant decoding in this analysis would indicate the neural pattern coding for the nontarget-probe distance resembles the neural pattern coding for the target-probe distance when the target was previously cued.

Relationship between probe processing and behavioural error

To test if target-probe / nontarget-probe distance decoding quality was related to behavioural performance on the orientation recall task, we fit a general linear model with EEG decoding values averaged over the time-window of interest (response-locked: –0.5-0s; probe-locked: 0-0.5s) as the predictor variable and absolute error as the outcome variable. The design matrix included a constant term. We fit the models separately for congruent and incongruent trials. We performed one-sample t-tests on the β weights to test whether the quality of decision-signal decoding was associated absolute error. We expected target-probe distance decoding quality to be associated with reduced error (one-tailed, <0) and nontarget-probe distance decoding quality to be associated with greater error (one-tailed, >0).

Code accessibility

Data and custom code will be made available on publication and can be made available to readers upon request.

Results

Orientation recall is faster and more precise on congruent relative to incongruent trials

Participants performed well on the incidental task with mean accuracy of .972 (SD=.021) and mean median RT on correct trials was .592 s (SD=.052). Correct incidental task trials were included in the subsequent analyses.

Behavioural orientation recall data are shown in Figure 1B-C. As expected, mean absolute error was significantly lower for congruent (M=13.75, SD=3.22) compared to incongruent trials (M=20.67, SD=6.08): t_{25}=-7.01, p<.001, d=-1.38 (Figure 1C, left panel). Similarly, RTs were significantly lower for congruent (M=.55, SD=.12) compared to incongruent trials (M=.67, SD=.15): t_{25}=-8.35, p<.001, d=-1.64 (Figure 1C, right panel). Selecting the cued colour in the

incidental task facilitated speed and accuracy of subsequent recall of other features belonging to the same item relative to the other item in memory.

[FIGURE 1]

Congruency affects all sources of error

Errors in orientation judgments may arise from multiple sources including reduced precision, reduced probability of reporting the correct memory item (target rate), increased likelihood of guessing (guess rate) and/or increased likelihood of reporting the other memory item (non-target rate). The estimated parameters of the swap model are shown in Figure 1D. Congruency affected all sources of error. Precision (K) was significantly higher for congruent (M=4.50, SD=1.30) than incongruent (M=3.27, SD=1.27) trials: t_{25}=6.31, p<.001, d=1.24. The same was true for the target rate (congruent: M=.94, SD=.06; incongruent: M=.79, SD=.14): t_{25}=5.74, p<.001, d=1.13. Guess rates were significantly lower for congruent (M=.05, SD=.05) than incongruent (M=.17 , SD=.13) trials: t_{25}=-4.73, p<.001, d=-.927, and so was the non-target rate (congruent: M=.02, SD =.02; incongruent: M=.04, SD=.06): t_{25}=-2.20, p=.019, d=-.431.

Non-predictive colour cue strengthens the neural representation of the cued item's orientation

If attentional selection of the cued colour for the incidental task spreads to other features belonging to the cued item, we would expect the neural representation of the cued item's orientation to be reinstated following the cue.

First, we ensured that the orientation of both items could be decoded during the WM array epoch (Figure 2A). Orientation decoding quality was significant for both memory items (cued: t_{25}=4.55, p<.001, d=.894; 0.07-0.76 s from WM array onset, cluster-corrected p<.001; uncued: t_{25}=5.31, p<.001, d=1.04; 0.06-0.80 s from WM array onset, cluster-corrected p<.001) and there was no significant difference between them (p=.780). We use labels 'cued' and 'uncued' to keep consistent across analyses even though the cue does not appear until later in the trial. Note that decoding quality dropped to near baseline prior to cue onset.

[FIGURE 2]

Our main analysis of interest tested whether decoding quality for the cued item's orientation re-emerged following cue onset (Figure 2B). As predicted, decoding quality for the cued item's orientation was significantly greater than zero in the cue epoch (t_{25}=1.93, p=.032, d=.379; 0.30-0.54 s after the cue, cluster-corrected p=.025), while the orientation of the uncued item could

not be decoded (p=.201). However, there was no significant difference in decoding quality between the cued and uncued item (p=.177). In line with our hypothesis, the results suggest that attentional selection of the cued colour strengthens the neural representation for the cued item's orientation, even though the item's orientation is irrelevant at the time of the cue.

In an exploratory analysis, we applied a complementary spatiotemporal decoding approach to boost power to detect neural patterns coding for memorised orientations using EEG (Figure 2C-D). Confirming the standard decoding results, the orientations of both items could be decoded following WM array onset (100-500 ms; cued: t_{25}=8.09, p<.001, d=1.59; uncued: t_{25}=8.01, p<.001, d=1.57), and there was no difference in decoding strength between the two items (p=.318).

In the cue-locked analysis of interest (100-500 ms) the orientation of the cued item, but not the uncued item, could be decoded (cued: t_{25}=3.56, p<.001, d=.698; uncued: p=.157) and decoding strength was significantly higher for the cued item's orientation compared to the uncued item (t_{25}=1.84, p=.039, d=.361). Thus, the results of the spatiotemporal decoding analysis broadly confirm the standard decoding results and further suggest that the cue-induced representational enhancement is greater for the cued than the uncued item, even though the cued item's orientation was no more relevant to the task than the uncued item's orientation.

The EEG-based orientation decoding results were not explained by fixational eye-movements (Mostert et al., 2018; Quax et al., 2019). We repeated the same orientation decoding analysis on eye-position data instead of EEG data (Figure 3A-B). Decoding from eye-position was not quite significant for the cued item's orientation following cue onset, but was significant for the uncued item's position (cued: t_{25}=1.63, p=.057, d=.321; uncued: t_{25}=1.81, p=.041, d=.355; difference: p=.351). However, two participants were outliers and showed strong eye-based orientation decoding for the cued item. Thus, we repeated the EEG decoding analyses excluding these participants to show that EEG-based orientation decoding was not driven by participants with strong decoding from eye-position (Figure 3C-D).

[FIGURE 3]

Cued orientation decoding quality predicts behavioural congruency effect

If the observed representational enhancement of the cued item's orientation strengthens the memory quality and accessibility of the cued item relative to the uncued item, decoding

quality for the cued item's orientation should predict the size of the behavioural congruency benefit (congruency–incongruent). We tested this prediction in a set of exploratory analyses.

The congruency benefit on reaction times showed a negative correlation with cued item orientation decoding quality (rs_{24}=–.424, p=.032; Figure 2E), indicating that greater enhancement of the cued item's orientation may improve memory accessibility for the cued relative to the uncued memory item. There was no significant correlation between cued orientation decoding quality and overall RT on congruent (p=.225) or incongruent trials (p=.656).

The congruency benefit on mean absolute recall error similarly showed a trend for a negative correlation with orientation decoding quality for the cued item following the cue, indicating that participants with greater orientation decoding quality showed smaller mean absolute error on congruent relative to incongruent trials, though this relationship did not reach significance (rs_{24}=–.388, p=.051; Figure 2E). We repeated the analysis separately for congruent and incongruent trials, to test whether orientation decoding quality predicted overall mean absolute error in each condition. There was no relationship between orientation decoding quality and mean absolute error on congruent trials. On incongruent trials, greater orientation decoding quality for the cued item was associated with higher error when the other item was probed for orientation recall, confirmed by a positive correlation (rs_{24}=.45, p=.022; Figure 2F).

We found no evidence for a trial-wise relationship between post-cue orientation decoding and behavioural measures on either congruent (RT: p=.752, absolute error: p=.825) or incongruent trials (RT: p=.788; absolute error: p= .612).

The results suggest that representational boost of the cued item's orientation strengthens the relative accessibility and quality of the cued item relative to the uncued item in memory.

Colour cue triggers alpha suppression contralateral to cued item's memorised location

Next, we asked if selecting the cued colour for the incidental task engaged spatial attention toward the original position of the cued item, even though location information was irrelevant to the task. If so, we expected to see alpha suppression contralateral to the cued item's original position. Figure 4A shows contralateral minus ipsilateral power across a range of frequencies following cue onset, and Figure 4B shows mean contra- minus ipsilateral alpha power (8-

14Hz). As predicted, contralateral alpha power was significantly lower than ipsilateral alpha power in the time-window following cue onset (0.2-0.8 s): t_{25}=-4.38, p<.001, d=-.877. Time-resolved alpha lateralization following cue onset (0-1 s) was significant from 0.2 to 0.71 s (cluster-corrected p<.001). Thus, colour-selection in WM is associated with transient suppression in contralateral alpha power, even when memory for location is not demanded by the task.

Colour cue triggers gaze bias toward cued item's memorised location

Eye-position data are shown in Figure 4C-D. As predicted, mean gaze position was significantly biased toward the original position of the cued item as indicated by towardness values greater than zero (M=.223%, SD=.529%) following cue onset (0.2-0.8s): t_{25}=2.15, p=.021, d=.423 (Figure 3D). Time-resolved towardness following incidental cue onset (0-1 s) was significant from 0.37 to 0.63 s (cluster-corrected p=.038), suggesting colour-selection triggers a transient shift in gaze position toward the cued item's memorised position. It is worth noting that, while statistically robust, this bias was tiny on average (.223% of 6° = 0.013° visual angle). As also evident from the 2D histogram (4C), it did not arise from overt saccades to the item's location on a proportion of trials, but rather small but consistent shifts of gaze position.

[FIGURE 4]

Contralateral alpha suppression predicts speed of orientation recall

Next, we tested whether there was a trial-wise relationship between measures of spatial attention (contralateral alpha suppression and gaze bias) and behavioural performance (error and RT) in a set of exploratory analyses. If spatial attention toward the cued item's original position is involved in the selection and prioritization of the cued item's other features in WM, contralateral alpha suppression and gaze bias should be associated with faster and/or more precise performance on congruent trials. If selection of the cued item comes at the expense of the uncued item in memory, alpha suppression and gaze bias might be associated with slowed and/or less precise performance on incongruent trials.

Trial-wise contralateral alpha suppression was associated with faster RTs on congruent trials (Figure 4E; t_{25}=2.20, p=.019, d=.432; 0.56-0.81 s, cluster-corrected p=.025). While ß weights were mostly negative on incongruent trials, contralateral alpha suppression did not significantly predict slowed RTs on those trials (p=.207). There was a significant difference in ßs between congruent and incongruent trials (t_{25}=2.18, p=.039, d=.257; 0.52 to 0.73 s cluster-

corrected p=.030). The results indicate that greater suppression of alpha contralateral to the cued item's original position was associated with faster RTs when the same item was subsequently probed for orientation recall. There was no strong evidence of a trial-wise relationship between contralateral alpha suppression and absolute error (congruent: p=.231; incongruent: t_{25}=-1.42, p=.084, d=-.278; difference: p=.202).

As for EEG-based orientation decoding, we tested whether mean contralateral alpha suppression predicted the size of the behavioural congruency effect on median RT and mean absolute error across participants. The congruency benefit on RT showed a trend for a positive correlation with alpha lateralization following the cue, as expected if participants with greater alpha lateralization show lower RT on congruent relative to incongruent trials (congruent–incongruent), but this relationship was not significant (r_{24}=.381, p=.056; Figure 4F). There was no significant correlation between the congruency benefit on mean absolute error and alpha lateralization, although the relationship was in the expected direction (r_{24}=.331, p=.099).

Together, the results show that greater contralateral alpha suppression following the colour cue predicts faster response initiation when the same item is subsequently probed, suggesting contralateral alpha suppression may improve the accessibility of the cued item in memory.

Gaze bias is associated with greater error on incongruent trials

The trial-wise regression analyses showed that gaze bias toward the cued item's original position was associated with greater absolute error on incongruent trials (t_{25}= 2.56, p=.008, d=.503; 0.42 to 0.98 s cluster-corrected p=.023; 1.10 to 1.5 s, cluster-corrected p = .029). There was no significant relationship with gaze bias and absolute error on congruent trials (p=.286) and ß weights were significantly greater for incongruent than congruent trials (t_{25}=-2.38, p=.025, d=-.467). There was also no evidence of a trial-wise relationship between gaze bias and RT (congruent: p=.225; incongruent: p=.656; difference: p=.798). The results indicate that the more participants' gaze was biased toward the cued item in memory, the higher the error when the other memory item was probed for orientation recall.

At the subject-wise level, there was a significant correlation between participants' mean towardness following the cue and the behavioural congruency effect on mean absolute error (r_{24}=-.404, p=.042), suggesting participants who showed greater gaze bias toward the cued item's original location showed smaller mean absolute orientation recall error on congruent

relative to incongruent trials. There was no significant correlation between mean towardness and the congruency benefit on RT (p=.309).

In summary, both alpha lateralization and gaze bias significantly influenced behaviour, but in distinct ways. While alpha lateralization primarily predicted reaction times (especially on congruent trials), gaze bias primarily predicted errors (especially on incongruent trials).

No correlation between alpha lateralization and gaze bias

There was no relationship between alpha lateralization and gaze bias following the cue (0.2-0.8s from cue onset) on a trial-wise (ß weights not significantly different from zero, p=.598) or participant-wise level (r_{S24}=-.264, p=.193).

Target-probe comparison signal stronger on congruent trials

The results so far indicate that attentional selection of the cued colour during the delay is associated with faster and more precise recall of the cued relative to the uncued item's orientation, enhances the neural representation of the cued item's orientation and triggers spatial attention toward the cued item's original position. Next, we investigated whether attentional selection of the cued item during the delay modulates how effectively the memory items are compared to the probe at recall in a set of exploratory analyses.

At recall, participants were required to compare the orientation of the randomly oriented probe stimulus to the probed item in memory (target) in order to turn the dial to the appropriate angle. The neural representation of the absolute distance between the random probe orientation and the target (target-probe distance) may signal how well the memory item is compared to the probe at recall. The absolute target-probe distance could be decoded prior to response onset (mean decoding -0.5 to 0 s) on both congruent (t_{25}=6.16, p<.001, d=1.21; -0.696 to 0 s, cluster-correction p <.001) and incongruent trials (t_{25}=2.95, p=.003, d=.579; -0.28 to 0 s, cluster-corrected p=.018; Figure 5A). Target-probe distance decoding quality was significantly higher on congruent than incongruent trials (t_{25}=3.41, p=.002, d=.668; -0.404 to 0 s, cluster-corrected p=.002). Thus, the target-probe distance comparison signal was stronger prior to response onset when the item had previously been selected from memory.

RTs varied as a function of probe-target distance, with slower RTs when the probe-target distance was small (Figure 5C). However, RT was not differently modulated by probe-target

distances on congruent and incongruent trials (Figure 5C; 2-way ANOVA, interaction between congruency and probe-target distance, $F_{7,25}$=1.31, p=.247), making it unlikely that the reported difference in probe-target distance decoding quality between congruent and incongruent trials is driven by differences in RT (or, equivalently, how recently the probe has appeared).

[FIGURE 5]

Target-probe comparison signal predicts orientation recall error

We tested whether there was a trial-wise relationship between the target-probe distance signal prior to response onset and behavioural performance (Figure 5B). We found that target-probe distance decoding quality correlated with orientation recall error on both congruent (t_{25}=-3.36, p=.001, d=-.660) and incongruent (t_{25}=-3.17, p=.002, d=-.622) trials, indicating that a stronger representation of the probe-target distance magnitude predicted lower behavioural error in both cases. There was no difference between conditions (p=.293).

Probe-locked decoding of target-probe and nontarget-probe distance

If the item selected for the incidental task during the delay is in a prioritised state for comparison with the probe at recall, we might expect a comparison signal between the probe and the *cued* item regardless of whether it is the target item probed for orientation recall, as is the case on congruent trials, or the nontarget item, as is the case on incongruent trials. This analysis focused on the probe-evoked response (0-0.5 s following probe onset). We would expect any erroneous comparison to the non-target to occur right after probe onset, especially since swap errors were very rare in both conditions (<5%), meaning that by the time participants start responding they have likely retrieved the target item.

The absolute distance between the probed memory item and the random probe orientation (target-probe distance) could be decoded following probe onset on congruent, but not on incongruent trials (congruent: t_{25}=3.91, p<.001, d=.768; 0.25 to 0.5 s, cluster-corrected p=.001; incongruent: p=.274; Figure 5D). Mirroring the results of the response-locked analysis, target-probe distance decoding strength was significantly higher on congruent than incongruent trials (t_{25}=2.92, p=.004, d=.573; 0.282 to 0.477 s, cluster-corrected p=.012).

By contrast, the absolute distance between the nontarget item in memory and the random probe orientation (nontarget-probe distance) could be decoded following probe onset on incongruent trials, but not on congruent trials (congruent: t_{25}=1.59, p=.062, d=.313; incongruent: t_{25}=2.92, p=.004, d=.572; 0.19 to 0.5 s, cluster-corrected p=.002). However, there was no significant difference between conditions (p=.212; Figure 4D).

If the cued item is incorrectly used as the template for comparison with the probe on incongruent trials, we might expect the neural pattern coding for the target-probe distance on congruent trials and the neural pattern coding for the nontarget-probe distance on incongruent trials to share overlapping representational formats. If so, a decoder trained on target-probe distance values on congruent trials should be able to decode the nontarget-probe distance on incongruent trials, which is what we found (t_{25}=3.21, p=.002, d=.630; 0.18 to 0.392 s, cluster-corrected p=.002; Figure 5E). That was not the case on congruent trials, where the nontarget had not been cued during the delay (t_{25}=1.58, p=.063, d=.310). However, nontarget-probe decoding quality was not significantly greater on incongruent than congruent trials (t_{25}=-1.17, p=.126, d=-.230).

In summary, following probe onset, the absolute distance between the probe and the *cued* memory item could be decoded regardless of whether the cued item is the target for orientation recall or not. The comparison signal with respect to the *uncued* item, on the other hand, could not be decoded immediately following probe onset, even when it was the target for orientation recall. These results indicate that the cued item may be prioritized for comparison with the probe at recall.

Representational overlap between cued item target-probe distance and nontarget-probe distance interferes with precision of orientation recall

We would expect the extent to which the nontarget is mistakenly used as a template for comparison with the probe to interfere with behavioural performance. If so, greater nontarget-probe distance decoding quality (when trained on congruent trials with *target*-probe distance labels) should predict worse orientation recall performance. This was the case on incongruent trials, where there was a positive trial-wise relationship between decoding quality and absolute error (ßs>0; t_{25}= 2.36, p=.013, d=.464). There was no evidence of a relationship on congruent trials (p=.898), where nontarget-probe decoding quality was not significant. ß

weights were significantly larger on incongruent than congruent trials (t_{25}=-2.18, p=.039, d=-.427). There was no evidence of a relationship with RT on congruent (p=.359) or incongruent (p=.625) trials.

These results were only found after training the decoder on the target-probe distance, indicating that interference was driven by erroneously treating the cued non-target as a target. By comparison, when the decoder was trained on the non-target probe distance, there was no significant relationship with error (congruent: p= .872; incongruent: p=.727) or RT (congruent: p=.385; incongruent: p=.369).

Discussion

Attention can be allocated to internal content in WM to flexibly select and prioritise task-relevant objects or features for future action. Here we asked whether attentional selection of a single object-feature during the delay enhances processing of other features belonging to the same item. We showed behavioural and neural evidence that attentional selection spreads between an object's features in WM. Exploratory analyses further revealed that attentional selection during the WM delay prioritises the selected object for subsequent decision-making.

Behaviourally, attentional selection of the cued colour for the incidental task was associated with faster and more precise recall of the same item's orientation, relative to the orientation of another memory item. This is consistent with previous studies showing benefits of non-predictive retro-cues in WM (Yi et al., 2008; Zokaei et al., 2014b, 2014a; Souza et al., 2015). The current results show that the behavioural benefit extends to other features of the cued object, even when location-based binding of object-features was not encouraged by the incidental task. Thus, attentional selection of the cued colour privileges memory for associated features belonging to the cued object. A potential alternative explanation for the observed congruency effect is that the cued colour becomes more accessible in memory, perhaps making it a better cue for subsequent retrieval. To directly assess the effect of colour selection on processing of other features in WM during the maintenance delay, we looked at neural and eye-tracking signatures of feature-based and spatial attention.

When WM content is prioritized via task-relevant predictive retro-cues, it is represented more strongly in neural activity patterns compared to unattended content (Lewis-Peacock et al.,

2012; LaRocque et al., 2013; Sprague et al., 2016; Wolff et al., 2017). Here, even though incidental cues were not predictive or relevant to the memory task, we found that the neural representation for the cued item's orientation as decoded from EEG voltages was enhanced following the cue, while the orientation of the uncued item could not be decoded. Orientation decoding quality across participants predicted the size of the behavioural congruency effect on RT and, to some extent, error, suggesting that neural representational enhancement boosts the accessibility and quality of the cued relative to the uncued item's orientation in WM. An exploratory spatiotemporal decoding analysis showed greater orientation decoding quality for the cued relative to the uncued item. This was the case even though the cued item's orientation was no more relevant to the task than the other orientation in WM.

Contralateral alpha suppression and gaze bias are thought to track the focus of spatial attention in WM (Poch et al., 2014; Myers et al., 2015b; Wallis et al., 2015; van Ede et al., 2019; Van Ede et al., 2020). In the present study, a nonpredictive, centrally presented colour cue was associated with alpha lateralization and gaze bias relative to the cued memory item's original position. These indicate a shift of covert attention to the item's previous location, possibly as part of the process of selecting the item in memory. Spatial attention may be involved in colour selection from WM, even when location is irrelevant, consistent with previous literature proposing that location may have a special status in WM (Treisman and Zhang, 2006; Pertzov and Husain, 2014; Poch et al., 2017; Schneegans and Bays, 2017; Cai et al., 2019). Even though location was irrelevant, it was still a distinguishing feature in our task. It is possible spatial attention is not similarly engaged when memory items share overlapping spatial locations and location information may interfere with performance (but see Pertzov & Husain, 2014).

The extent to which participants directed spatial attention to the cued colour's original position predicted orientation recall performance, suggesting spatial attention plays a functional role in WM maintenance and retrieval, even of non-spatial features (Cai et al., 2019). Even though contralateral alpha suppression and gaze bias are both thought to track allocation of spatial attention in WM, the two measures did not correlate at the trial-wise or participant-wise level and they showed distinct correlations with behaviour. Contralateral alpha suppression following cue onset predicted faster orientation recall on congruent trials, suggesting alpha plays a role in selecting the cued content, making it more accessible for retrieval. Gaze bias relative to the cued item's position was associated with greater error on

incongruent trials, suggesting gaze bias toward the cued item may negatively affect memory quality for the other item. Future studies should further disentangle the relationship between alpha lateralization and gaze bias as measures of spatial attention and the ways in which spatial attention facilitates WM for non-spatial features.

The results discussed so far build a case for obligatory spread of attention between an object's features in WM. Selecting a feature of a WM object enhances processing of other features belonging to the cued item, even when enhancing them is irrelevant to the memory task. Object-based selection in WM can be contrasted with the alternative hypothesis that it is possible to selectively attend to a single object feature of a multi-feature object encoded into WM. Several studies have shown that WM selectively encodes and represents object-features that are relevant to the current task (Woodman and Vogel, 2008; Serences et al., 2009; Bocincova and Johnson, 2019). Behavioural benefits afforded by feature-based retro-cues further support feature-based selection in WM (Niklaus et al., 2017; Hajonides et al., 2020). While the abovementioned studies initially appear at odds with the current findings, predictive retro-cues may allow irrelevant features to be dropped from memory, thereby freeing up WM resources to process relevant features (Souza and Oberauer, 2016). In the present experiment, the non-predictive cue did not allow removal of any features. The object may therefore constitute a relevant "bundle" of connected features that are enhanced whenever a single feature of that bundle is selected (Brady et al., 2011), consistent with theoretical models of attention in WM that retrieve features through auto-associative pattern completion (Lansner et al., 2013; Manohar et al., 2019).

Analyses of neural processing of the probe at recall showed that a comparison signal was automatically computed for the cued object in WM, boosting orientation recall when the cued item was also the target (congruent), but interfering with retrieval when it was not (incongruent). Recent theoretical perspectives predict that attentional selection may prioritise selected information for interaction with new sensory input (Olivers et al., 2011; Myers et al., 2017; Heuer et al., 2020; Olivers and Roelfsema, 2020). For example, currently relevant WM objects automatically guide visual attention, while items that are currently irrelevant but might become relevant later do not (Peters et al., 2009). Filtering new sensory input through a neural pattern coding for the prioritised memory item (i.e., a matched-filter) may serve as an efficient mechanism for computing the relevant decision signal—in this case, the angular

distance of the mnemonic template relative to the probe (Sugase-Miyamoto et al., 2008; Hayden and Gallant, 2013; Myers et al., 2015a; Muhle-Karbe et al., 2021).

The cued item may be maintained in a functionally active state, that influences ongoing processing, by virtue of its recent use in decision-making for the incidental task (Myers et al., 2017; Heuer et al., 2020; Olivers and Roelfsema, 2020; Stokes et al., 2020). Such a state could be established through strengthened functional connectivity in the network coding for the cued item (Rerko and Oberauer, 2013; Oberauer and Lin, 2017; Manohar et al., 2019; Olivers and Roelfsema, 2020) and enhanced links to neural areas responsible for configuration of task-sets, motor planning and decision-making (Myers et al., 2017; Olivers and Roelfsema, 2020). This may facilitate faster and more efficient comparison between the probe and the cued item, providing a potential mechanism for the congruency effect observed in this study.

In conclusion, attentional selection of a single feature during the delay enhanced processing of other features belonging to the same object. The results show that well-known object-based attention mechanisms exist for internal attention and highlight objects as a key unit of selection in WM. Attentional selection may prioritise a WM object for comparison with new sensory input, even when the selected object has no special relevance to subsequent behaviour, providing a potential mechanism for non-predictive cueing benefits in WM.

Citation gender diversity statement

We assessed the gender balance of papers referenced in this article using a combination of the Citation-Transparency Chrome extension to google scholar and manual gender estimation from author's google scholar profiles. Of the 63 papers with named authors, 57.1 % (n=36) had male first author and male last author, 20.6% (n=13) had a female first author and a male last author, 19 % (n=12) had a male first author and a female last author and 3.17% (n=2) had a female first author and a female last author.

References

American Clinical Neurophysiology Society (1991) American Electroencephalographic Society guidelines for standard electrode position nomenclature. J Clin Neurophysiol 8:200–202.

Baddeley AD (2003) Working memory: Looking back and looking forward. Nat Rev Neurosci

4:829–839.

Baddeley AD (2012) Working memory: Theories, models, and controversies. Annu Rev Psychol 63:1–29.

Bays PM, Catalao RFG, Husain M (2009) The precision of visual working memory is set by allocation of a shared resource. J Vis 9:7.1-11.

Bays PM, Taylor R (2018) A neural model of retrospective attention in visual working memory. Cogn Psychol 100:43–52.

Bocincova A, Johnson JS (2019) The time course of encoding and maintenance of task-relevant versus irrelevant object features in working memory. Cortex 111:196–209.

Brady TF, Konkle T, Alvarez GA (2011) A review of visual memory capacity: Beyond individual items and toward structured representations. J Vis 11:4–4.

Brainard DH (1997) The Psychophysics Toolbox. Spat Vis 10:433–436.

Cai Y, Sheldon AD, Yu Q, Postle BR (2019) Overlapping and distinct contributions of stimulus location and of spatial context to nonspatial visual short-term memory. J Neurophysiol 121:1222–1231.

Chun MM, Johnson MK (2011) Memory: Enduring traces of perceptual and reflective attention. Neuron 72:520–535.

Desimone R, Duncan J (1995) Neural Mechanisms of Selective Visual Attention. Annu Rev Neurosci 18:193–222.

Duncan J (1984) Selective attention and the organization of visual information. J Exp Psychol Gen 113:501–517.

Engbert R, Mergenthaler K (2006) Microsaccades are triggered by low retinal image slip. Proc Natl Acad Sci U S A 103:7192–7197.

Ernst ZR, Boynton GM, Jazayeri M (2013) The spread of attention across features of a surface. J Neurophysiol 110:2426–2439.

Gazzaley A, Nobre AC (2012) Top-down modulation: Bridging selective attention and working memory. Trends Cogn Sci 16:129–135.

Griffin IC, Nobre AC (2003) Orienting attention to locations in internal representations. J Cogn Neurosci 15:1176–1194.

Grootswagers T, Wardle SG, Carlson TA (2016) Decoding Dynamic Brain Patterns from Evoked Responses: A Tutorial on Multivariate Pattern Analysis Applied to Time Series Neuroimaging Data. Cogn Neurosci J 57:1–21.

Hajonides JE, van Ede F, Stokes MG, Nobre AC (2020) Comparing the prioritization of items and feature-dimensions in visual working memory. J Vis 20:25.

Hayden BY, Gallant JL (2013) Working memory and decision processes in visual area v4. Front Neurosci 7:18.

Heuer A, Ohl S, Rolfs M (2020) Memory for action: a functional view of selection in visual working memory. Vis cogn 28:388–400.

Kiyonaga A, Egner T (2013) Working memory as internal attention: Toward an integrative account of internal and external selection processes. Psychon Bull Rev 20:228–242.

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

Lansner A, Marklund P, Sikström S, Nilsson L-G (2013) Reactivation in Working Memory: An Attractor Network Model of Free Recall. PLoS One 8:e73776.

LaRocque JJ, Lewis-Peacock JA, Drysdale AT, Oberauer K, Postle BR (2013) Decoding Attended Information in Short-term Memory: An EEG Study. J Cogn Neurosci 25:127–142.

Ledoit O, Wolf M (2004) A well-conditioned estimator for large-dimensional covariance matrices. J Multivar Anal 88:365–411.

Lepsien J, Nobre AC (2007) Attentional modulation of object representations in working memory. Cereb Cortex 17:2072–2083.

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

Manohar SG, Zokaei N, Fallon SJ, Vogels TP, Husain M (2019) Neural mechanisms of attending to items in working memory. Neurosci Biobehav Rev 101:1–12.

Mostert P, Albers AM, Brinkman L, Todorova L, Kok P, de Lange FP (2018) Eye Movement-Related Confounds in Neural Decoding of Visual Working Memory Representations. eneuro 5:ENEURO.0401-17.2018.

Muhle-Karbe PS, Myers NE, Stokes MG (2021) A Hierarchy of Functional States in Working Memory. J Neurosci 41:4461 LP – 4475.

Myers NE, Rohenkohl G, Wyart V, Woolrich MW, Nobre AC, Stokes MG (2015a) Testing sensory evidence against mnemonic templates. Elife 4:1–25.

Myers NE, Stokes MG, Nobre AC (2017) Prioritizing Information during Working Memory: Beyond Sustained Internal Attention. Trends Cogn Sci 21:449–461.

Myers NE, Walther L, Wallis G, Stokes MG, Nobre AC (2015b) Temporal Dynamics of Attention during Encoding versus Maintenance of Working Memory: Complementary Views from Event-related Potentials and Alpha-band Oscillations. J Cogn Neurosci 27:492–508.

Niklaus M, Nobre AC, Van Ede F (2017) Feature-based attentional weighting and spreading in visual working memory. Sci Rep 7:1–10.

O'Craven KM, Downing PE, Kanwisher N (1999) fMRI evidence for objects as the units of attentional selection. Nature 401:584–587.

Oberauer K, Lin H-Y (2017) An Interference Model of Visual Working Memory. Psychol Rev 124:21–59.

Olivers CNL, Peters J, Houtkamp R, Roelfsema PR (2011) Different states in visual working memory: When it guides attention and when it does not. Trends Cogn Sci 15:327–334.

Olivers CNL, Roelfsema PR (2020) Attention for action in visual working memory. Cortex.

Oostenveld R, Fries P, Maris E, Schoffelen JM (2011) FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput Intell Neurosci 2011:1–9.

Panichello MF, Buschman TJ (2021) Shared mechanisms underlie the control of working memory and attention. Nature 592:601–605.

Pertzov Y, Husain M (2014) The privileged role of location in visual working memory. Attention, Perception, Psychophys 76:1914–1924.

Peters JC, Goebel R, Roelfsema PR (2009) Remembered but Unused: The Accessory Items in Working Memory that Do Not Guide Attention. J Cogn Neurosci 21:1081–1091.

Poch C, Campo P, Barnes GR (2014) Modulation of alpha and gamma oscillations related to retrospectively orienting attention within working memory. Eur J Neurosci 40:2399–2405.

Poch C, Capilla A, Hinojosa JA, Campo P (2017) Selection within working memory based on a color retro-cue modulates alpha oscillations. Neuropsychologia 106:133–137.

Quax SC, Dijkstra N, van Staveren MJ, Bosch SE, van Gerven MAJ (2019) Eye movements explain decodability during perception and cued attention in MEG. Neuroimage.

Rerko L, Oberauer K (2013) Focused, unfocused, and defocused information in working memory. J Exp Psychol Learn Mem Cogn 39:1075–1096.

Schneegans S, Bays PM (2017) Neural Architecture for Feature Binding in Visual Working Memory. J Neurosci 37:3913–3925.

Serences JT, Ester EF, Vogel EK, Awh E (2009) Stimulus-specific delay activity in human primary visual cortex. Psychol Sci 20:207–214.

Souza AS, Oberauer K (2016) In search of the focus of attention in working memory: 13 years of the retro-cue effect. Attention, Perception, Psychophys 78:1–22.

Souza AS, Rerko L, Oberauer K (2015) Refreshing memory traces: Thinking of an item improves retrieval from visual working memory. Ann N Y Acad Sci 1339:20–31.

Sprague TC, Ester EF, Serences JT (2016) Restoring Latent Visual Working Memory Representations in Human Cortex. Neuron 91:694–707.

Stokes MG, Muhle-Karbe PS, Myers NE (2020) Theoretical distinction between functional states in working memory and their corresponding neural states. Vis cogn 28:420–432.

Sugase-Miyamoto Y, Liu Z, Wiener MC, Optican LM, Richmond BJ (2008) Short-term memory trace in rapidly adapting synapses of inferior temporal cortex. PLoS Comput Biol 4.

The MathWorks Inc. (2017) MATLAB and Statistics Toolbox Release 2017b.

Treisman AM, Zhang W (2006) Location and binding in visual working memory. Mem Cognit 34:1704–1719.

Van Ede F, Board AG, Nobre AC (2020) Goal-directed and stimulus-driven selection of internal representations. Proc Natl Acad Sci U S A 117:24590–24598.

van Ede F, Chekroud SR, Nobre AC (2019) Human gaze tracks attentional focusing in memorized visual space. Nat Hum Behav 3:462–470.

Wallis G, Stokes MG, Cousijn H, Woolrich MG, Nobre AC (2015) Frontoparietal and cingulo-opercular networks play dissociable roles in control of working memory. J Cogn Neurosci 27:2019–2034.

Wolff MJ, Jochim J, Akyürek EG, Buschman TJ, Stokes MG (2020) Drifting codes within a stable coding scheme for working memory Tong F, ed. PLOS Biol 18:e3000625.

Wolff MJ, Jochim J, Akyürek EG, Stokes MG (2017) Dynamic hidden states underlying working-memory-guided behavior. Nat Neurosci 20:864–871.

Woodman GF, Vogel EK (2008) Selective storage and maintenance of an object's features in visual working memory. Psychon Bull Rev 15:223–229.

Yi DJ, Turk-Browne NB, Chun MM, Johnson MK (2008) When a thought equals a look: Refreshing enhances perceptual memory. J Cogn Neurosci 20:1371–1380.

Zhang W, Luck SJ (2008) Discrete fixed-resolution representations in visual working memory. Nature 453:233–235.

Zokaei N, Manohar S, Husain M, Feredoes E (2014a) Causal Evidence for a Privileged Working Memory State in Early Visual Cortex. J Neurosci 34:158–162.

Zokaei N, Ning S, Manohar S, Feredoes E, Husain M (2014b) Flexibility of representational states in working memory. Front Hum Neurosci 8:1–12.

Figures

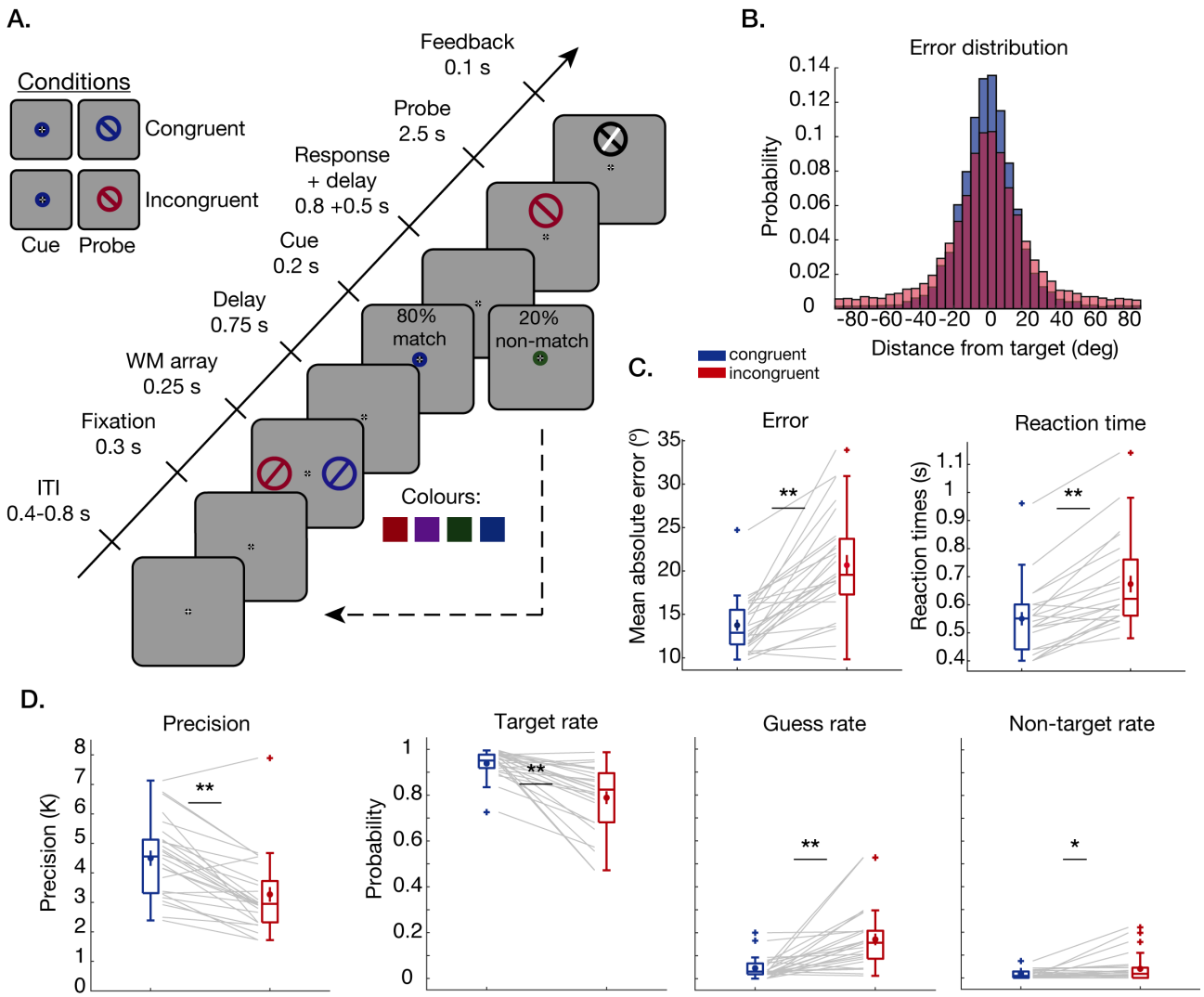

Figure 1. Task design and behavioural results. **A.** Task design. Participants remembered two oriented, coloured bars on each trial, presented laterally. During the maintenance delay, they were shown a non-predictive colour cue that asked them to make a speeded mismatch-response (i.e., does it differ from the colours in memory?; 'incidental task'). At the end of the trial, they reported the orientation of one of the items, cued by colour. This item was either the same as was also cued on the incidental task (congruent condition, 50%), or the other item (incongruent condition, 50%). **B.** Histogram of angular error (deg) relative to target orientation across all participants for congruent (blue) and incongruent (red) trials. **C.** Left panel: boxplot of mean absolute error. Right panel: boxplot of median reaction times (time until response initiation). Overlaid mean and SEM. Grey lines show individual participants. **D.** Swap model parameters. Boxplots of precision (K), target rate, guess rate and non-target rate. *p<.05, **p<.001.

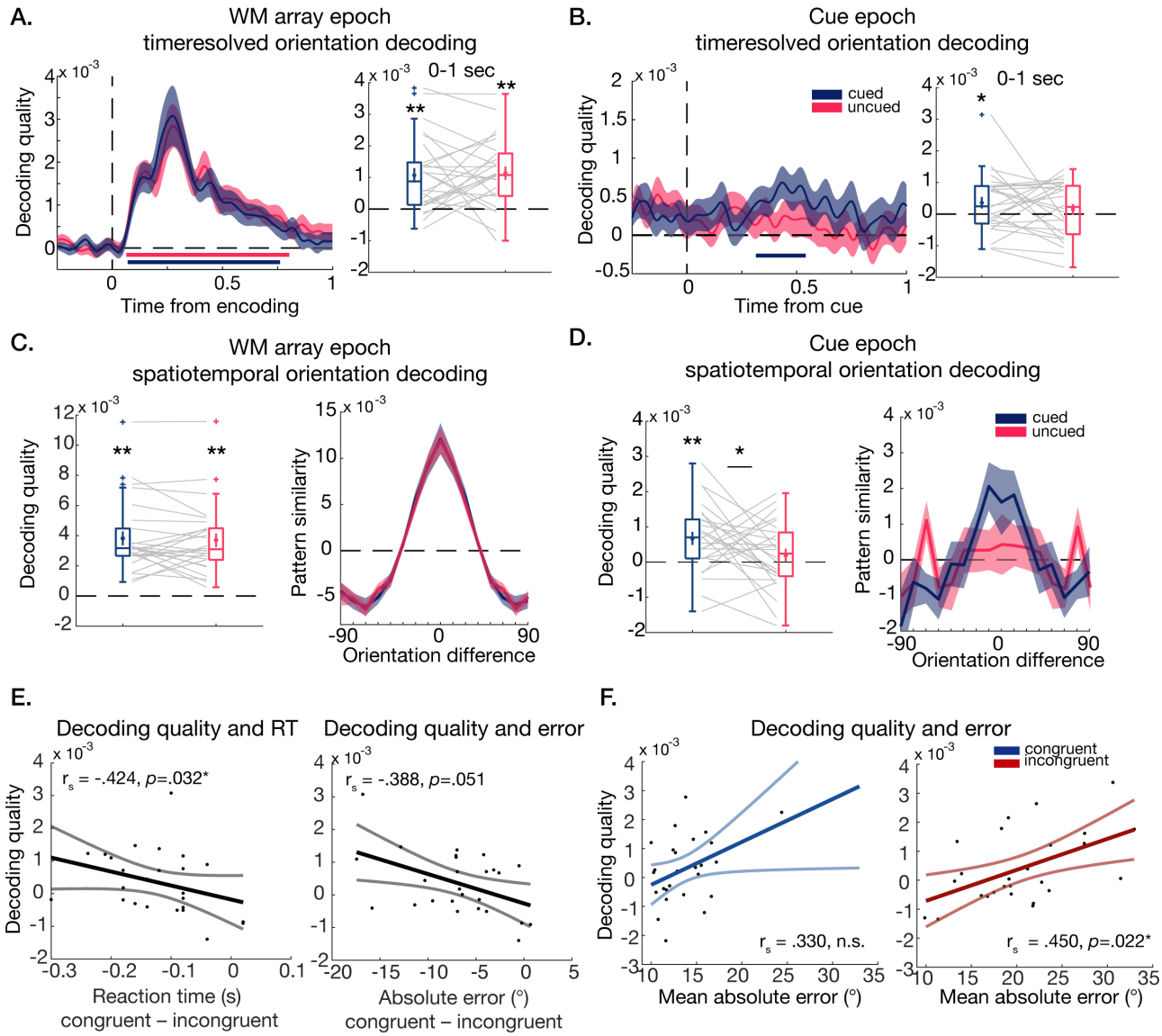

Figure 2. Decoding orientation from EEG voltages. **A.** Left: Decoding quality following WM array onset. Shaded area shows SEM. Bars show clusters of significant decoding. Boxplot shows mean decoding across the encoding epoch (0-1s from WM array onset). Overlaid mean and SEM. Grey lines show individual participants. **B.** Same as A but for cue epoch. **C.** Spatiotemporal orientation decoding. Left: Boxplot of decoding quality following WM array onset (100-500 ms). Right: Mahalanobis distance pattern similarity as a function of orientation difference. Shaded area shows SEM. **D.** Same as C but following cue onset (100-500 ms). **E.** Scatter plot of relationship between mean cued orientation decoding quality and congruency effect (congruent – incongruent) on RT (left) and mean absolute error (right). Linear regression line and 95% confidence intervals overlaid. **F.** Scatter plot of relationship between mean cued orientation decoding quality and mean absolute error across participants for congruent (left) and incongruent (right) trials. *p<.05, **p<.001.

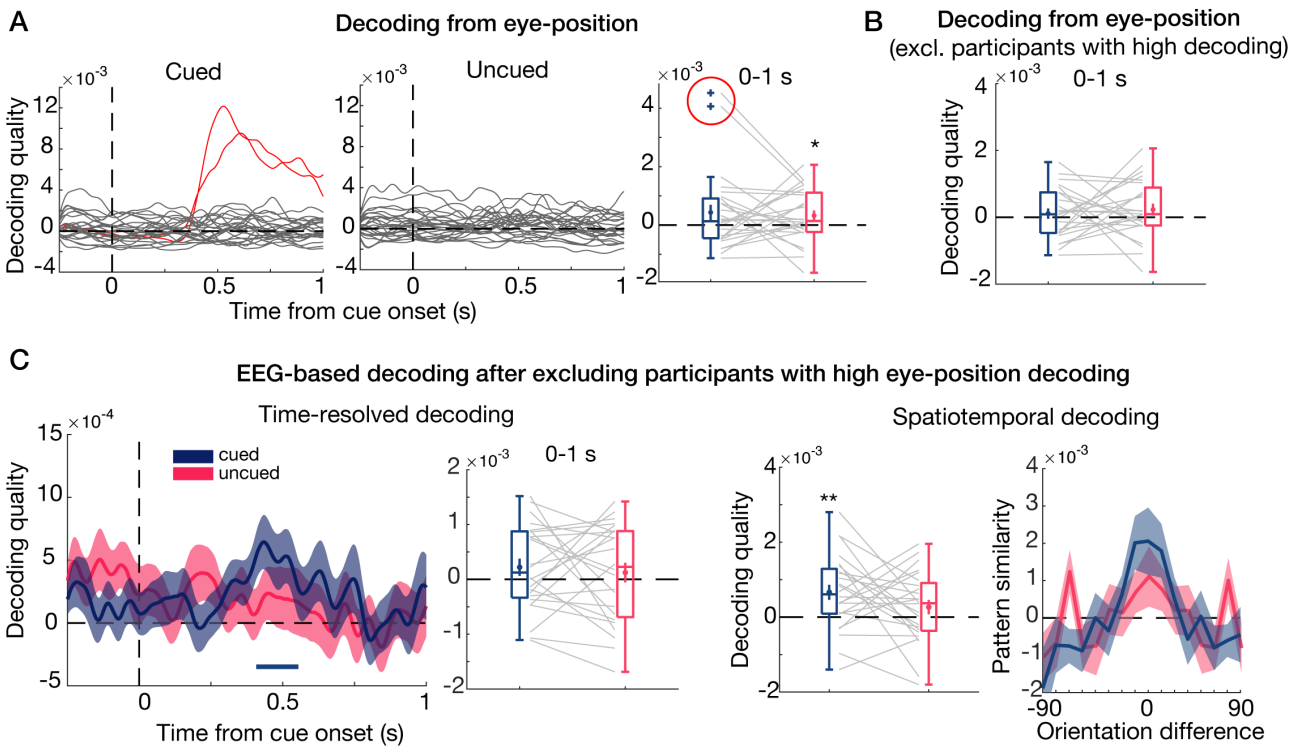

Figure 3. Controlling for fixational eye-movements. **A.** Eye-position based decoding of orientation. Decoding quality over time for all participants (individual lines) for cued (left) and uncued (right) item. Boxplot shows mean decoding quality across 0-1s from cue onset. Overlaid mean and SEM. Grey lines show individual participants. Red lines and circle represent two outliers excluded from control analyses. **B.** Boxplot showing mean eye-position based orientation decoding after excluding participants with high eye-position decoding. **C.** EEG-based decoding of orientation after excluding participants with high eye-position decoding. Left: standard time-resolved decoding quality. Blue bar represents significant time-points after cluster-correction (0.38-0.53 s., p=.053). Mean decoding (cued: t_{23}=1.40, p=.088, d=.285; uncued: p=.252; difference: p=.317). Right: spatiotemporal decoding quality (cued: t_{23}=3.23, p=.001, d=.659; uncued: t_{23}=1.37, p=.092, d=.280; difference: t_{22}=1.44, p=.082, d=.293). *p<.05, **p<.001.

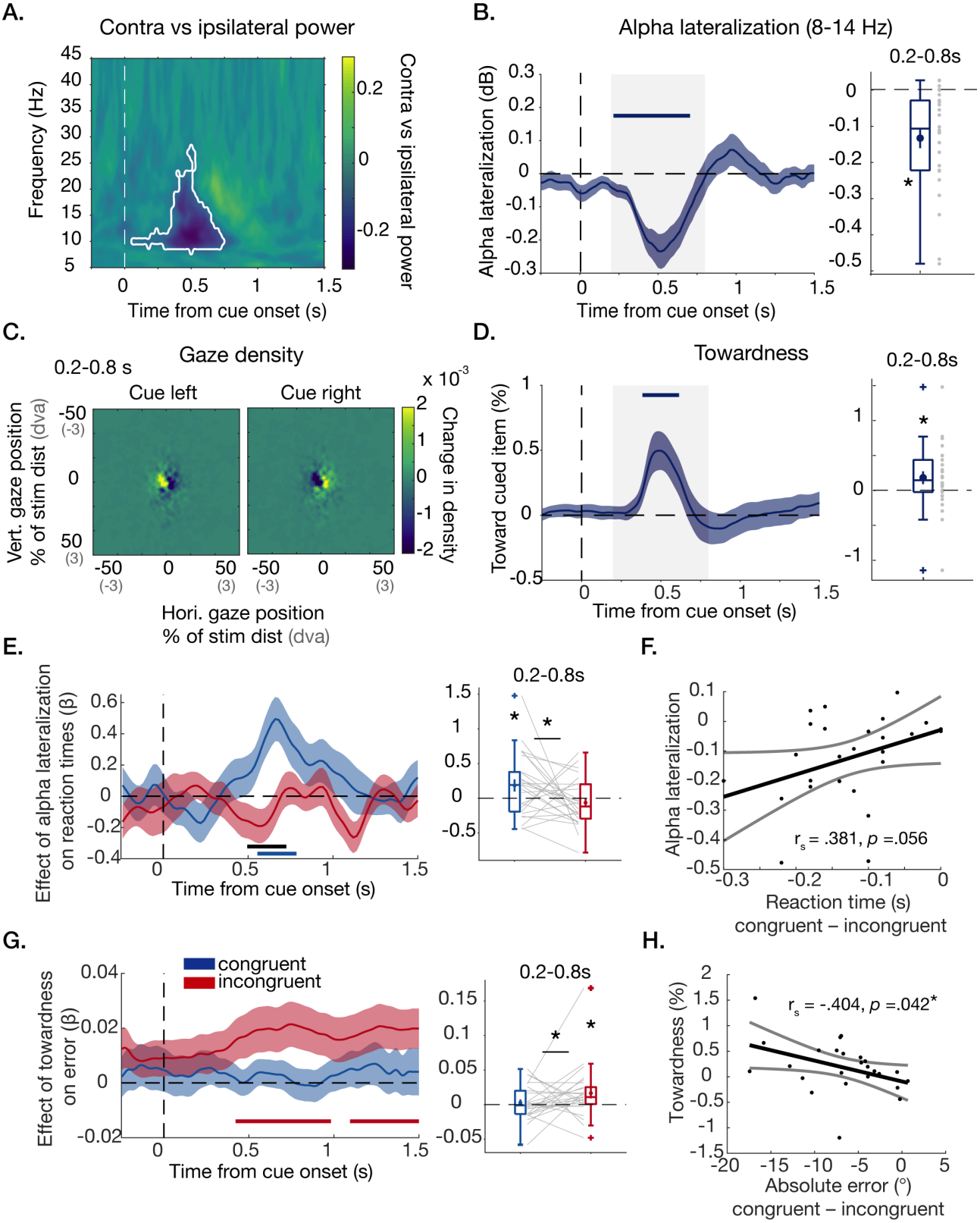

Figure 4. Alpha lateralization and gaze bias **A.** Contra minus ipsilateral power relative to original position of cued item over time from cue onset. White contours show cluster of significant contralateral suppression. **B.** Left: alpha lateralization (contra minus ipsilateral alpha power; 8-14 Hz) over time from cue onset. Grey shaded area indicates time window of interest (0.2-0.9 s). Shaded area shows SEM. Right: Boxplot of mean alpha lateralization across cue epoch (0.2-0.8 s from cue onset). Mean and

SEM overlaid. Grey dots show individual participants. **C.** Two-dimensional histogram of gaze density for cue left and cue right trials (subtracting mean gaze across conditions). **D.** Mean horizontal gaze position for right cue trials minus the mean gaze position for left cue trials, divided by two ("towardness") over time from cue onset. Positive values indicate gaze bias toward cued item. Right: Boxplot of mean towardness across cue epoch (0.2-0.8 s from cue onset). **E.** Relationship between alpha lateralization and reaction times. Regression weights (β) from trial-wise regression with alpha lateralization and reaction times for congruent (blue) and incongruent (red) trials. Left: β weights from time-resolved regression. Right: β weights from regression with mean alpha lateralization across cue epoch. Blue bar shows non-zero beta-weights for congruent trials and black bar shows significant difference between congruent and incongruent trials. **F.** Scatterplot of relationship between mean alpha lateralization and congruency effect (congruent–incongruent) on RT across participants. Linear regression line and 95% confidence intervals overlaid. **G-H.** Same as E-F for relationship between gaze bias and absolute error. *p<.05, **p<.001.

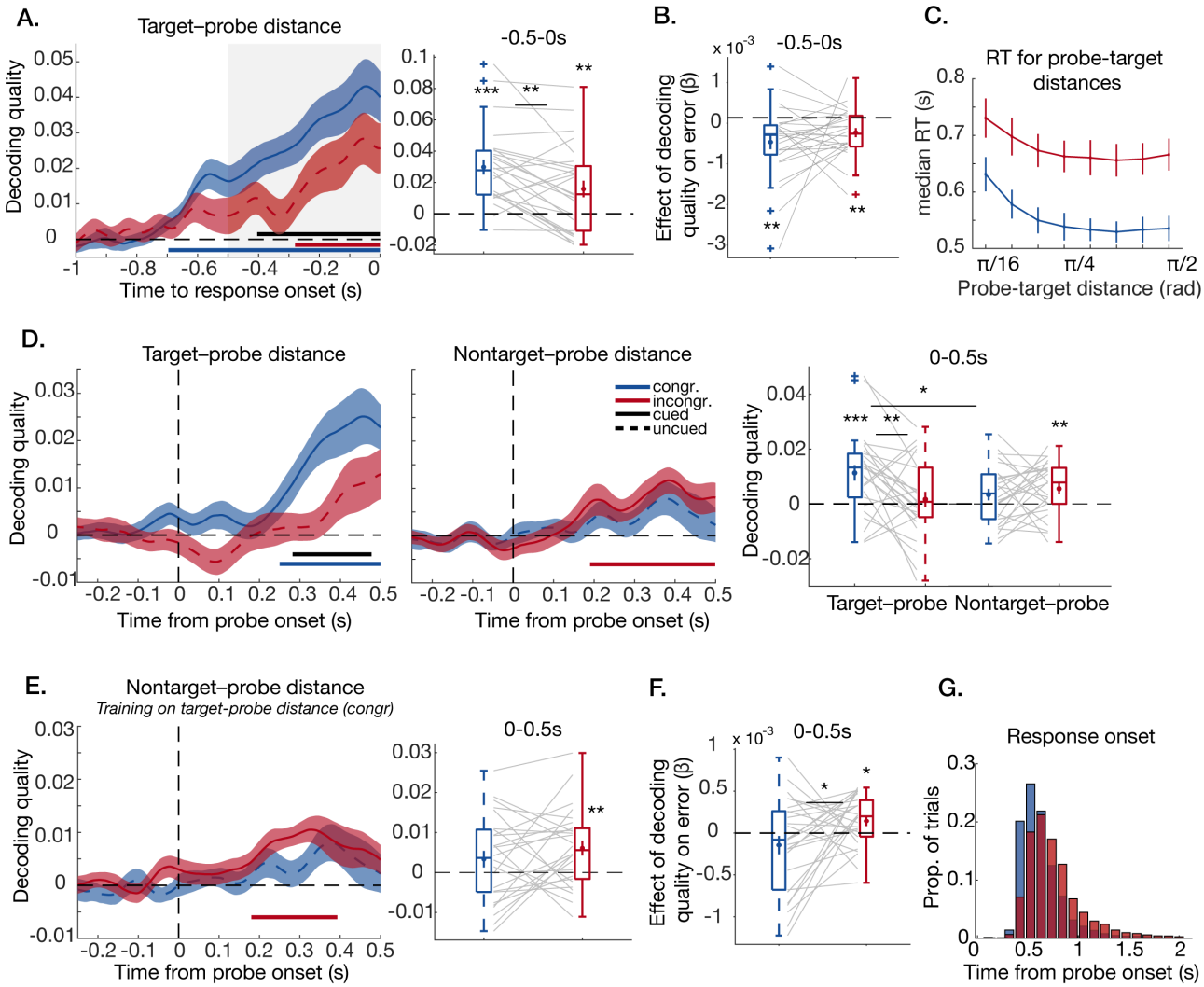

Figure 5. Decoding target-probe and nontarget-probe distance from EEG voltages. **A.** Decoding target-probe distance (absolute distance between random probe orientation and orientation of probed item in memory) for congruent (blue) and incongruent (red) trials time-locked to response onset. Shaded area shows SEM. Bars show significant decoding. Grey shaded area indicates time window of interest (-0.5–0s). Boxplot (right) shows mean decoding strength before response onset (-0.5-0 s). Mean and SEM overlaid. Grey lines show individual participants. **B.** Relationship between target-probe distance

decoding and orientation recall error. Regression weights (β) from trial-wise regression with mean probe-target distance decoding (-0.5-0s before response onset) and absolute error for congruent (blue) and incongruent (red) trials. **C.** Mean median reaction time (s) as a function of probe-target distance (in radians) on congruent (blue) and incongruent (red) trials. Error bars show SEM. **D.** Decoding of target-probe distance (left) and nontarget-probe distance (middle) for congruent (blue) and incongruent (red) trials time-locked to probe onset. Boxplot (right) shows mean decoding strength following probe onset (0-0.5s). Solid lines show the item that was cued during the delay and dashed lines show the uncued item. **E.** Same as middle panel in D, except decoding training data was sampled from congruent trials labelled with target-probe distance values. **F.** Relationship between nontarget-probe distance decoding (trained on congruent target-probe distance; shown in E) and orientation recall error. Regression weights (β) from trial-wise regression with mean nontarget-probe distance decoding (0-0.5 after probe onset) and absolute error. **G.** Histogram of response onset relative to probe onset across participants for congruent (blue) and incongruent (red) trials. *p<.05, **p<.01, ***p<.001.