Maintenance of colour memoranda in activity-quiescent working memory states: Evidence from impulse perturbation

Güven Kandemir[1,2]*, Sophia A. Wilhelm[1], Nikolai Axmacher[3], & Elkan G. Akyürek[1]

[1] Department of Experimental Psychology, University of Groningen, The Netherlands

[2] Department of Experimental and Applied Psychology, Vrije Universiteit Amsterdam, The Netherlands

[3] Department of Neuropsychology, Faculty of Psychology, Ruhr University Bochum, Germany

Running head: Impulse perturbation of colours in working memory

27/06/2023

Word count: 8929

Address correspondence to:

Güven Kandemir

Institute for Brain and Behavior, Vrije Universiteit Amsterdam

Van der Boechorststraat 7, 1081 BT Amsterdam, The Netherlands

Email: g.kandemir@vu.nl

Telephone: +31 (0) 205981055

1

Abstract

The neural mechanisms underlying working memory maintenance pose a challenge for investigation, as sustained neural activity may not always be observable. To address this, the method of impulse perturbation has been employed to examine memorized information during activity-quiescent periods. However, this approach has mainly focused on spatially localized or referenced stimuli, leaving it unclear whether non-spatial memoranda share similar neural maintenance mechanisms. This study aimed to fill this gap by applying the impulse perturbation method to working memory for colours, which are inherently non-spatial stimuli. EEG data from 30 participants performing a delayed match-to-sample task were analysed, with one of the presented items being retro-cued as task-relevant. Our findings indicate that both cued and uncued colours could be decoded from impulse-evoked activity, in contrast to previous reports on working memory for orientation gratings. Additionally, we explored colour decoding from ongoing oscillations in the alpha band and discovered that cued items could be decoded, potentially influenced by attention, whereas uncued items could not. These results suggest subtle differences between the representation of colours and stimuli with spatial properties. However, they also demonstrate that both types of information can be accessed through visual impulse perturbation, regardless of their specific neural states.

Key words: working memory; EEG; decoding; activity-silent representations; impulse perturbation

Introduction

Working memory is a system of components that enables the maintenance of information in an accessible state in the absence of sensory stimulation (Cowan, 2017). The neural basis of working memory has been of increasing interest in recent years. One particularly striking outcome has been that although working memory maintenance has been traditionally associated with sustained neural activity during the memory delay period (Kojima & Goldman-Rakic, 1982; Goldman-Rakic,1995; Constantinidis et al., 2018), a number of studies failed to observe such sustained activity (LaRocque, et al., 2013; Lewis-Peacock, et al., 2012; Shafi et al., 2007; Watanabe & Funahashi, 2014). It was proposed that during this period without an observable neural correlate, memoranda could be retained in activity-silent states (Stokes, 2015; Wolff et al., 2017), which may rely on synaptic plasticity facilitated by elevated post-excitatory calcium and neurotransmitter levels (Buonomano & Maass, 2009; Barak & Tsodyks, 2014; Kozachkov et al., 2022; Lundqvist et al., 2018; Mongillo et al., 2008; Miller et al., 2018).

Wolff and colleagues have previously shown that the presentation of a high-contrast, standardized, but task-irrelevant stimulus during the memory delay may allow decoding of such activity-silent, or at least activity-quiescent, memory items that could not be detected from raw ongoing EEG (Wolff et al., 2017). While it is not yet exactly clear how this so-called impulse signal reveals the memory trace at a physiological level, Wolff and colleagues explained this using the analogy of sonar: The impulse allows measuring a 'hidden' state by attributing differences in the response to a stable stimulus to underlying differences in the network (Wolff et al., 2017). In our case, the sensory processing of an impulse signal is thought to perturb the initially 'hidden' memory network, generating activity that can be measured with EEG, from which the state of this network can be inferred (Stokes, 2015; Wolff et al., 2015; 2017; 2020a; 2020b). Studies using the impulse perturbation approach have reported successful results for memories of orientations (Wolff et al., 2015; 2017; 2020a; 2020b), numerosity (Fornaciai & Park, 2020), and auditory tone frequencies and sequences (Fan et al., 2021; Fan & Luo, 2022; Wolff et al., 2020b).

Although impulse perturbation has thus been used with different stimuli, it is yet uncertain whether the representational patterns observed to date will hold universally for different kinds of content that are

3

maintained in working memory. In particular, one crucial similarity between previously tested memoranda is that they allow a transformation of the task relevant information into spatial coordinates (which may also aid memory). For example, each member of a set of orientations can be represented as different points on the edge of an imaginary circle around a fixed point on a plane surface. In a similar fashion, higher or lower numerosity and tone frequency can be easily converted to different elevation levels laid out on a similar surface (e.g., a high tone may be visualized high on the vertical axis). This possibility of representing information with spatial position leaves open the question of whether the neural signature of the maintenance of this kind of information also applies to that of non-spatial stimulus attributes.

The reliance on spatial properties in previous studies also brings confounding risks with it. For example, it might involve the deployment of spatial attention, which might affect the alpha band of the EEG in particular (Worden et al., 2000). Bae and Luck (2018) investigated the contribution of spatial attention to the decoding of orientation items in working memory. They found that while alpha band activity only conveyed information about the (attended) location of their stimuli (but see also Barbosa et al., 2021), and while decoding of ongoing EEG during the delay period mainly provided significant information about the item-specific orientation, it also reflected its location. Furthermore, differences that exist between stimuli or conditions in terms of spatial attention come with the risk that (voluntary or involuntary) eye movements may follow suit. Neural data can be confounded by correlating eye movements and gaze fixations, especially under active viewing conditions (Nishimoto et al., 2017; Thielen et al., 2018; 2019). Activity in the brain, particularly in earlier visual regions, may reflect viewpoint-specific, retinotopically-organized information that will vary considerably when the eyes move around. Such activity cannot be easily discerned from other aspects of sensory and cognitive processing, including memory maintenance (Mostert et al., 2018; Quax, et al., 2019).

Considering the theoretical limitations and potential confounding issues with regard to the exclusive reliance on spatial properties in previous impulse-based experiments on working memory maintenance, we set out in the current study to overcome these by using intrinsically non-spatial stimuli, namely colours. Colours can be decoded successfully from fMRI (Brouwer & Heeger, 2009; Persichetti et

4

al., 2015), MEG (Hermann, et al., 2022; Rosenthal et al., 2021; Sandhaeger et al., 2019), and EEG data (Bocincova & Johnson, 2019; Hajonides et al., 2021; Sandhaeger et al., 2019). Furthermore, the colour-space is similar to orientation space with regard to parametric differences between distinct equiluminant colours (Hajonides et al., 2021). Thus, colours seem a suitable non-spatial substitute to extend earlier orientation-based pinging studies (Wolff et al, 2015; 2017; 2020a; 2020b). Apart from thus changing the memory items, we also presented them serially, at the same location, rather than lateralized, as was originally done, and rotated the colour wheel that served as the response probe on each trial, thus removing all spatial aspects from the original task.

We conducted a set of pre-registered analyses, based on those reported in the original paper by Wolff and colleagues (2017), and added decoding analyses of alpha band activity. Our results show that trial-specific colour information could be successfully predicted from the activity evoked by the visual impulse, similar to orientation gratings in earlier studies (Wolff, 2015; 2017; 2020a; 2020b). In addition to the task-relevant cued colour, dynamic impulse-driven activity also revealed the uncued colour. Conversely, while alpha power decoding yielded a sustained trace of the cued memory item, this was not the case for the uncued colour. These findings extend previous work on orientation decoding, and suggest that different features might elicit (slightly) different maintenance mechanisms. The present outcomes also highlight that impulse-driven decoding can reveal memoranda in distinct memory states, independent of the allocation of spatial attention.

Method

Participants

Thirty volunteers (23 female, M_{age} = 24.2, $Range_{age}$ = 20-36) that were recruited via social media adverts participated in this study in return for monetary rewards. Participant selection relied on the successful completion of a pre-screening test, which was a shortened version of the main experiment (288 trials). The preselection cut-off criterion was ≤30° of error in at least 70 % of trials. None of the volunteers were eliminated by the pre-screening. The sample size was based on earlier studies with similar designs (e.g., Wolff et al, 2017). All participants were informed about the experimental procedures as well as the

data sharing procedures, and written consent was obtained. The study was conducted in accordance with the Declaration of Helsinki (2008), and it was approved by the Ethical Committee of the Behavioural and Social Sciences Faculty of the University of Groningen (Study ID = PSY-1920-S-0385).

Apparatus and Stimuli

The experiment took place in a well-lit chamber where participants were seated 60 cm away from a 17" Samsung 797DF CRT monitor. The refresh rate was set to 100 Hz and the resolution was 1024 by 768 pixels. All stimuli were created and presented with the freely available Psychtoolbox 3 extension for Matlab (Brainard, 1997; Kleiner, et al., 2007).

The memory items and the probe consisted of coloured disks with a visual angle of 6.69°, which were presented in the centre of the screen. Their colours were randomly drawn from RGB conversions of 48 equiluminant colours equally distanced on the CIELAB colour-wheel, which were extracted from the freely available Matlab extension MemToolbox (Suchow et al., 2013). A grey background (RGB = 128, 128, 128) was maintained throughout the experiment. A black fixation dot with a white outline (0.25° of visual angle) was displayed in the centre of the screen at all times except during the presentation of the cue. The cue was a number, "1" or "2", indicating the serial position of the task-relevant ("cued") item in that trial, presented in Arial font in the centre of the screen (0.5° visual angle). The impulse was a large white disk, displayed in the centre of the screen with a visual angle of 13.38°. The response screen contained the probe in the centre of the screen and a colour-wheel surrounding the probe, which had a diameter of 10.05° and a width of 0.55° visual angle. A white line reaching from the centre to the edge of the colour circle indicated the momentary probe colour on the colour-wheel. In each trial, the colour-wheel was randomly rotated and a random colour was assigned to the probe disk. Responses were collected with an Xbox controller. Following a response, a happy or a sad smiley face was presented at the centre in Arial font, indicating accuracy (i.e., whether the absolute error was less or more than 30°).

Procedure

The experiment consisted of 1536 trials, which were completed in four consecutive sessions that were separated by breaks. Participants determined the duration of the breaks. In each session participants

6

completed 24 blocks, and after each block an average score per block was presented as feedback. Participants could start each block by pressing the SPACE bar, after which trials continued automatically until all the trials in the block were completed. At the beginning of each block, a "Get Ready" warning was presented first, after which the trials commenced.

Each trial started with the presentation of a fixation dot, which was shown for 700 ms on a grey background. This was followed by the serial presentation of two colours for 200 ms duration, each followed by a 900 ms delay. The cue was presented next for 200 ms, and a delay of 900 ms followed it. Next, the impulse signal was presented for 100 ms, and a consecutive delay was on display for 500 ms. The response screen was displayed next and stayed on the screen until a response was submitted. When the response screen was on display, participants could move the left stick on the controller to rotate the black bar presented on the probe, and change the colour of the probe circle. Once the desired colour was selected, the response was submitted by pressing X on the controller. The response was followed by a delay of 150 ms after which feedback was presented for 300 ms. The next trial began automatically after a random jitter with a range of 500 to 800 ms.
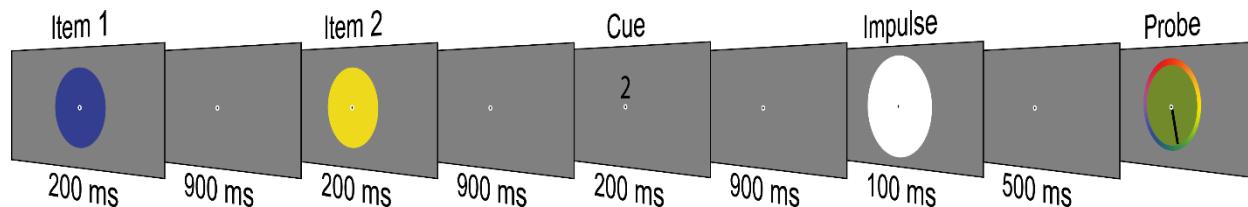

Figure 1. Overview of a single trial in the experiment. In each trial, the colours of two sequentially presented disks would be memorized. A retro-cue then indicated the temporal position of the task-relevant memory item. Finally, the randomly-oriented probe was rotated by the participants with a joystick to report the task-relevant colour. The numbers below the stimulus displays reflect their duration in ms.

EEG Acquisition and Pre-processing

The EEG was recorded with Brainvision Recorder software, and a TMSI Refa 8-64/72 amplifier using 62 Ag/AgCl sintered electrodes, which were placed according to the international 10-20 system. The data were recorded in reference to the average of all electrodes at a sampling rate of 1000 Hz. The ground

7

electrode was placed on the sternum, and eye movements were tracked via bipolar electrooculography with vertical electrodes above and below the left eye, and two horizontal electrodes on the ipsilateral sides of both eyes. The resistance at all electrodes was kept below 7 kΩ throughout the experiment.

Filtering and preprocessing were handled via the Matlab extensions Fieldtrip (Oostenveld et al., 2011) and EEGLAB (Delorme & Makeig, 2004). The data were re-referenced offline to the average of both mastoids. For the multivariate analyses on voltage values, the data were downsampled to 500 Hz, and filtered at 0.1 Hz high-pass and 40 Hz low-pass. Alpha power amplitudes were acquired by bandpass filtering the EEG signal with a 8 Hz high-pass and 12 Hz low-pass, by applying the Matlab function presented below to the data on each channel:

abs(hilbert(eegfilt(data, sample_rate, low_pass, high_pass))),

where low_pass and high_pass corresponded to the 8 and 12 Hz filters applied to our EEG data, and sample_rate corresponded to the 500 Hz sampling rate. The filter output was Hilbert transformed and the absolute of the product was calculated to get the real values of the transformation.

The voltage data and alpha power amplitudes were separately epoched to the onset of the memory items, the cue, and the impulse, covering a range starting from -150 ms relative to their onset until the onset of the next stimulus, thus forming distinct epochs for item 1, item 2, cue and impulse. Semi-automatic artefact rejection was completed by marking trials with high voltage variations and then inspecting all trials visually for channel drifts, muscle and eye artefacts. Drifting channels were interpolated using the spherical head model, whereas epochs with other artefacts and blinks were excluded from the analyses. In total 13.24% of the Item 1 epochs, 10.7% of the Item 2 epochs, 10.76% of the Cue epochs, and 9.99% of the impulse epochs were excluded.

Multivariate Analyses

Unless stated otherwise, all analyses were pre-registered at https://osf.io/uvxe7/. The decoding analyses were restricted to the 17 posterior channels (P7, P5, P3, P1, Pz, P2, P4, P6, P8, PO7, PO3, POz, PO4, PO8, O1, Oz and O2), replicating earlier studies that investigated visual working memory by means of impulse perturbation (Wolff et al., 2017; 2020a; 2020b).

Analysing the time window of interest

The primary analysis aimed to investigate the accuracy of trial-specific colour decoding, which was calculated using the data within a time window of interest. This time window of interest covered 100-400 ms relative to the presentation of a stimulus (e.g., memory item or impulse). This window was based on earlier studies, in which it was applied to capture the bulk of the EEG response to the eliciting stimulus, and in particular the dynamic response to the impulse (Wolff et al., 2015; 2017). First, the data were baselined by subtracting the average activity within the aforementioned period, which was done separately for each trial and electrode. Next, the data were downsampled by calculating the moving average over a 10 ms window. These downsampled values were then pooled over all 17 posterior electrodes, which yielded a single spatio-temporal pattern for each trial.

The trials were assigned to the closest one of 16 equidistant bins that covered the pre-determined colour-space, and were re-labelled with the centre value of that bin. This was repeated three times to cover the pre-determined colour-space (from 0° to 337.5°, 7.5° to 345°, and 15° to 352°, each in steps of 22.5°), thus providing 16 different colour conditions in each of the three runs. Next, the trials were partitioned into 8 folds with seven folds serving as the training set. Within the training set, the number of trials in each colour condition were equalized by subsampling the data. Subsequently, the data within each colour condition were averaged, forming 16 condition-specific spatio-temporal patterns. These spatio-temporal patterns of colour conditions were then convolved with a half cosine basis set raised to the 15th power to reduce noise and pool information across similar colours. The similarity between test trials and the averaged training data was quantified in Mahalanobis distances (De Maesschalck et al., 2000), yielding 16 distance values for each test trial. The covariance matrix was estimated from the entire training set by using a shrinkage estimator (Ledoit & Wolf, 2004). The distance values were mean-centred and sign-reversed, so that positive values indicated higher similarity. Finally, the values were convolved with a cosine-similarity function of the colour space (i.e., on the colour wheel). The product was the trial-specific decoding accuracy. The procedure was repeated 100 times with random folds and random sub-sampling, in order to

9

avoid sampling biases. The reported decoding accuracy for each participant was calculated by averaging all the products of all repetitions and all trials to get a single value per participant.

Time-course analyses

The time-course of the memory-related dynamic signal was investigated by sliding a 100 ms time-window across the epoch. In each step, data within the window were baselined by subtracting the mean activity within, and then the residual activity was downsampled to 100 Hz. The data were then pooled over electrode space, yielding the spatio-temporal pattern at that time point. The time-course analysis was highly similar to the analysis of the time window of interest. The trials were first re-labelled in accordance to the colour-space. The trials were divided into 8 folds with stratified sampling so that all conditions had an approximately equal representation in each fold. The training was set formed by 7 folds and the trials in these folds were distributed across 16 bins. The number of trials in each bin was equalized with random subsampling in order to avoid bias, and the data in each bin was averaged to form temporal patterns for each colour bin at each time point. The measure of similarity between the test trials and the 16 averaged colour patterns was calculated at each time point in Mahalanobis distances. A shrinkage estimator was used to estimate the covariance matrix from the training set (Ledoit & Wolf, 2004). The distance measures at each time point were reverse-signed and mean-centred, and then scaled by cosine convolution of the modelled colour space. The output was the decoding accuracy at each time point. The procedure was performed 3 times, once for each colour-space, and repeated 100 times to avoid sampling bias. The output was averaged over colour spaces, repetitions and trials for each participant.

Post-hoc, we conducted another time-course analysis on (absolute change in) alpha power, which was not included in the preregistration. The analysis was identical to the above, but applied to filtered 8-12 Hz EEG data, which was baselined over -200 to 0 ms relative to cue and impulse onset. The data were also downsampled to 125 Hz to save computational time. Finally, we also conducted the same time-course analysis on the EOG data to ensure the absence of any spatial correlation.

Representational similarity analyses

Representational similarity analyses (RSA) (Kriegskorte et al., 2008) was used to parametrically assess the relationship between colours during memory encoding and during the memory delay following the impulse. For the RSA, the EEG data from the same pre-determined 17 posterior electrodes within the time window (100 to 400 ms relative to a stimulus or an impulse signal onset) were prepared as in the decoding analysis of the time window of interest. The trials were re-labelled according to the first colour-space ($0° – 337.5°$ on the colour wheel) and grouped into 16 bins. The number of trials in each bin were equalized and 16 spatio-temporal patterns were generated by averaging trials in each bin. The pairwise differences between all bins were calculated in Mahalanobis distances to form the representational dissimilarity matrix (RDM). The covariance matrix was estimated from the training set with the use of a shrinkage estimator (Ledoit & Wolf, 2004). The entire procedure was repeated 100 times to account for selection biases and the final RDM was averaged over all repetitions for each participant.

The RDM was tested by the linear regression of two models. The first model covered the circular nature of 16 colour bins centred according to the first colour space. In other words, this model specifically assessed the hue differences in terms of angular space on the colour wheel. The second model was based on the behavioural output, which showed that the responses clustered around three primary colours. The second model thus assessed if the differences in neural space for the 16 colour bins could be explained by the differences between the three frequently reported discrete colours (green-blue, red and yellow). Both models were first converted to z-scores. Next, the RDM of each participant was regressed against each model separately. The diagonal segment for the RDM and models was excluded from analyses. The standardized slopes (beta values) were taken as an indication of the fit. The mean beta values were contrasted against zero by using a group permutation test.

Statistical assessment

Trial-wise decoding accuracy was averaged for each participant, yielding a single output (at each time point) per participant. The average decoding accuracy over all participants was contrasted against a null distribution (one-sided). In order to form this null distribution, the sign of the mean decoding accuracy (per participant) was flipped 100,000 times with 50 % probability, and the resultant mean was taken. The

11

proportion of the null distribution larger than the observed mean accuracy was calculated as the p value, which was labelled as significant if it was smaller than 5% ($p < 0.05$). For the time-course analysis, an additional cluster correction was applied to control for multiple comparisons, where the cut-off was set to 0.05 ($p < 0.05$). Confidence intervals were built by bootstrapping the mean ($n_{perm} = 100,000$). When decoding accuracies for two conditions were contrasted, the difference of these two conditions was calculated for each participant, and this group vector was tested against zero by using the same sort of permutation test as explained above. The RSA was similarly assessed by means of a group permutation test ($n_{perm} = 100,000$), which was applied on the beta values of the participants, reflecting the model fit to the pairwise differences reported in the RDMs.

Data availability

The data and analysis scripts of this study are publicly available at https://osf.io/bxmt8.

Results

Behavioural Results

The overall behavioural performance in the experiment was good; the mean error was 18.3° with a standard deviation of 32.5° (Figure 2A), relative to the centre of the bin to which the cued item belonged (adjusted error). Figure 2B presents the behavioural reports of the cued item, with the uniform colour bins overlaid. This plot reveals that despite the uniform distribution of the colours presented during the experiment, the reported colours seemed clustered around the primary hues, supporting earlier evidence that memory performance varies for different colours (Bae et al., 2014).

The influence of the uncued memory item on the report of the cued item was also investigated. Since only the cued item was reported in each trial, the presence of the uncued item was assessed as the effect of the similarity between the cued and the uncued colour on the degree of error. To take into consideration possible individual differences in colour perception (Emery & Webster, 2019), errors were first median-normalized within each colour bin. The normalized error values were then binned again as a function of the difference between the task-relevant cued item and the task-irrelevant uncued item (binwidth = 22.5°, moving window in steps of 7.5°), and the mean error was calculated within each bin. A permutation test

12

was applied with cluster correction to assess the deviation of the mean from zero at each unit of angular difference between the items. The report error was significantly different from zero for three difference ranges (Figure 2C, $M_{\text{Adjusted error}} \neq 0$, for cued – uncued difference, ranging from -161° to -3.75°, $p < 0.001$, from 3.75° to 116°, $p < 0.001$, and from 124° to 161°, $p = 0.024$). The cued item report errors deviated away from the uncued item, reflecting a repulsion away from the task-irrelevant colour.
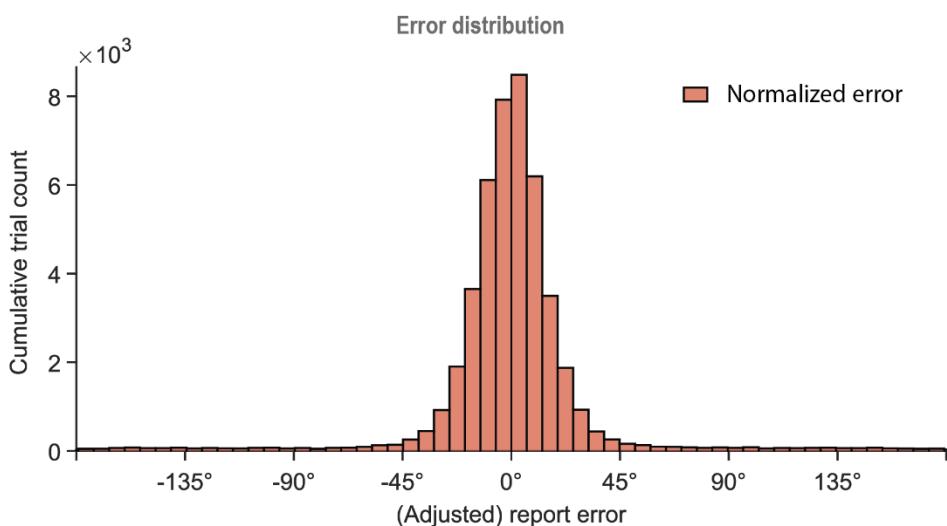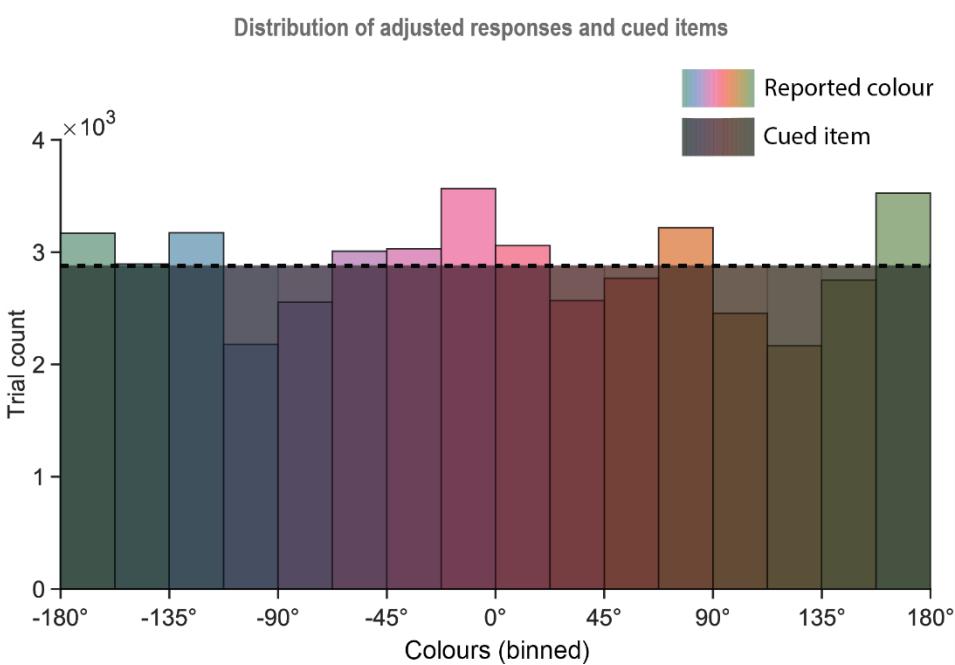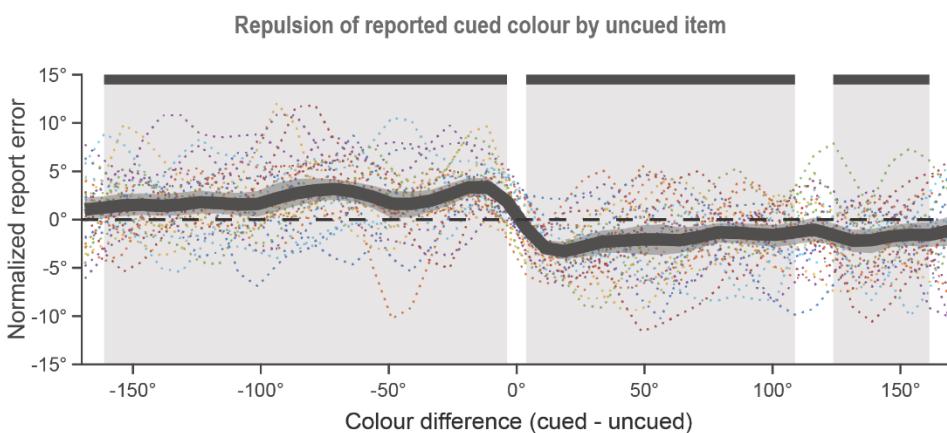
Figure 2. Behavioural plots. **(A)** Histogram of report errors relative to the task-relevant memory item. **(B)** Distribution of the cued colour bins (shaded, dashed line) and the reported colour bins, as a function of the angular values of the colour-space. **(C)** Normalized error as a function of the difference between the cued and the uncued item. The moving mean of the report error is calculated for a 22.5° wide bin in steps of 7.5°. The shaded area and the solid bars at the top mark the colour differences for which the adjusted error differed from 0°, according to the cluster-corrected permutation test ($p < 0.05$).

Decoding the time window of interest

Both the trial-specific colour of item 1 (Figure 3A, left, red, $p < 0.001$, one-tailed), and of item 2 (Figure 3A, left, blue, $p < 0.001$, one-tailed), could be successfully decoded after presentation. Interestingly, item 1 was also decodable within the critical period following the presentation of item 2 (Figure 3A, left, item 1 (2), red, $p < 0.001$, one-tailed). Decoding of item 1 after the onset of item 2 was nevertheless significantly lower than stimulus-driven decoding (difference $_{\text{item 1 – item1 (2)}}$, $p < 0.001$, one-tailed). The associated tuning curves (Figure 3A, right) reflected a parametric relationship between colours, in line with earlier reports (Hajonides et al., 2021).

After impulse presentation, not only the task-relevant, cued colour was decodable (Figure 3B, left, green, $p < 0.001$, one-tailed), but also the task-irrelevant uncued item (Figure 3B, left, grey, $p = 0.008$, one-tailed), in contrast to earlier studies that used orientation stimuli (e.g., Wolff et al., 2017). Although both cued and uncued memories could be traced, there was a clear difference in the strength of their representations (difference $_{\text{cued – uncued}}$, $p < 0.001$, one-tailed). Both the cued and uncued item showed parametric pattern similarity, as expected (Figure 3B, right).
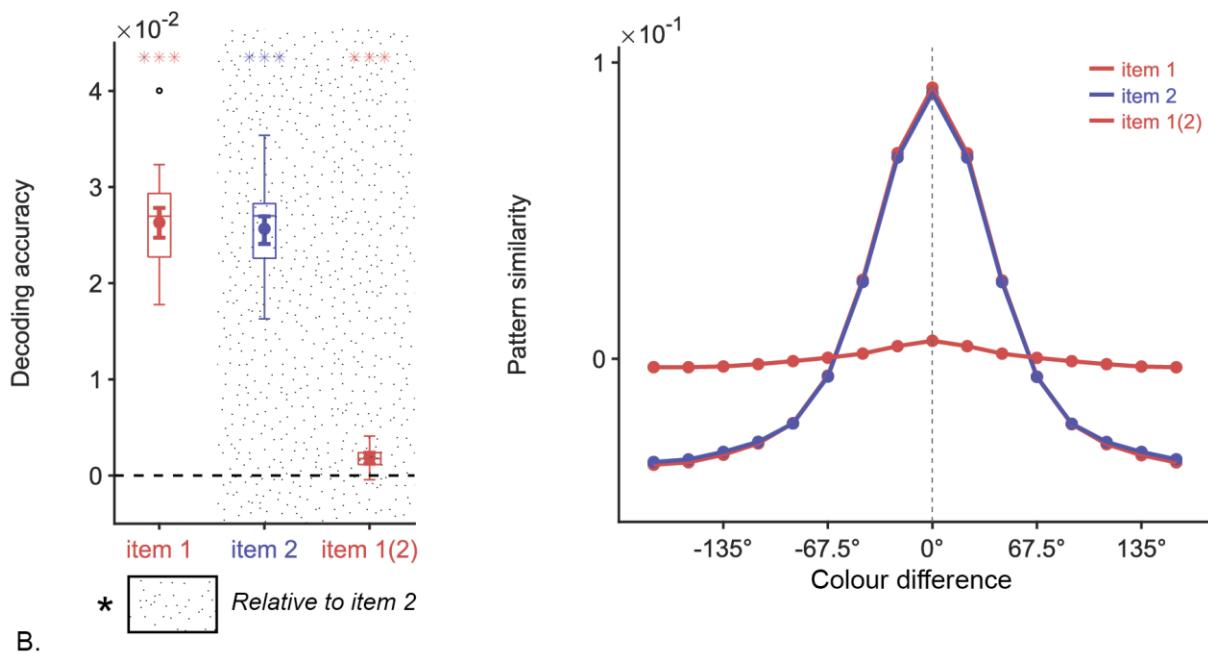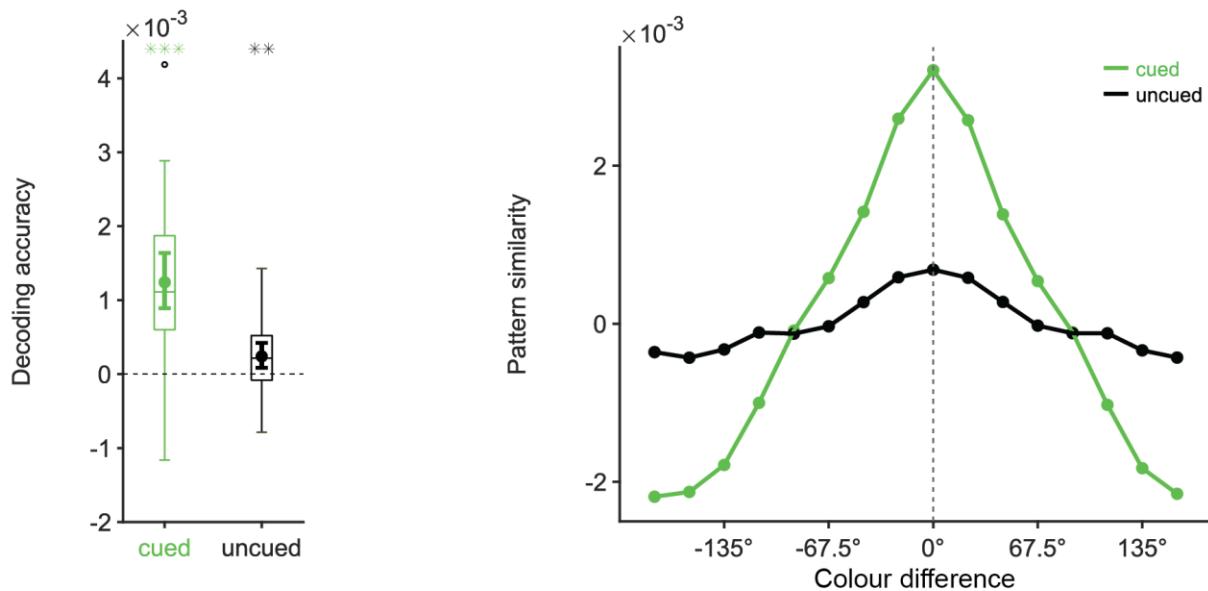
Figure 3. Boxplots and tuning curves showing decoding accuracy and pattern similarity (in arbitrary units) for memory items within the 100-400 ms time window of interest relative to the onset of item 1 and 2 presentation (A), as well as impulse onset (B). **(A)** Item 1 (red) could be successfully decoded after its onset, and after onset of item 2 ("item 1(2)"). Item 2 (blue) was also successfully decoded after its onset.

16

The pattern similarity plots (A, right) show a circular, parametric relationship between colours, reflected by the gradual change in pattern similarity as a function of the colour difference for item 1 and item 2. **(B)** Cued (green) and uncued (grey) items were both successfully decoded from the time window of interest, after impulse onset. The associated pattern similarity plots show that the parametric relationship between memory items was preserved (B, right). The boxes border the 25th and 75th percentiles, with the whiskers around the box stretching to 1.5 interquartile range. Asterisks indicate beta values that were significantly above zero (* , $p < 0.05$; ** , $p < 0.01$; *** , $p < 0.001$).

Time-course decoding

In addition to classic time-course decoding of colours, we additionally applied this analysis to alpha power. Motivating this addition was the recent suggestion that ongoing activity in the alpha band could reflect memory content in a sustained fashion, which questions the need to use the impulse perturbation method (Barbosa et al., 2021). The analyses were focused on the cue and impulse epochs, where the selection between task-relevant (cued) and task-irrelevant (uncued) items had been made. Furthermore, we were able to confirm that eye movements, attentionally driven or otherwise, did not affect decoding of the memory items by applying the same analysis to the EOG data (Supplementary Information).

The memory items could not be decoded from the voltage data following the presentation of the cue (Figure 4A, B). Conversely, both the cued and the uncued items were successfully decoded from alpha power in an earlier phase after the presentation of the cue (Figure 4C, green, 396 ms to 588 ms, $p_{cued} = 0.015$, two-tailed, corrected; Figure 4C, black, 364 ms to 516 ms, $p_{uncued} = 0.014$, two-tailed, corrected). Subsequently, in a later phase, only the task-relevant, cued colour was decodable in the alpha band without interruption for the remainder of the epoch until the onset of the impulse signal (Figure 4C, green, 700 ms to 1050 ms, $p_{cued} < 0.001$, two-tailed, corrected). The difference in decoding accuracy between cued and uncued items was also statistically significant for some time within this period, from 892 ms to 1028 ms relative to impulse onset ($p_{cued - uncued} = 0.037$, one-tailed, corrected). Across these phases, parametric pattern similarity was also apparent for both items (Figure 4D).

17
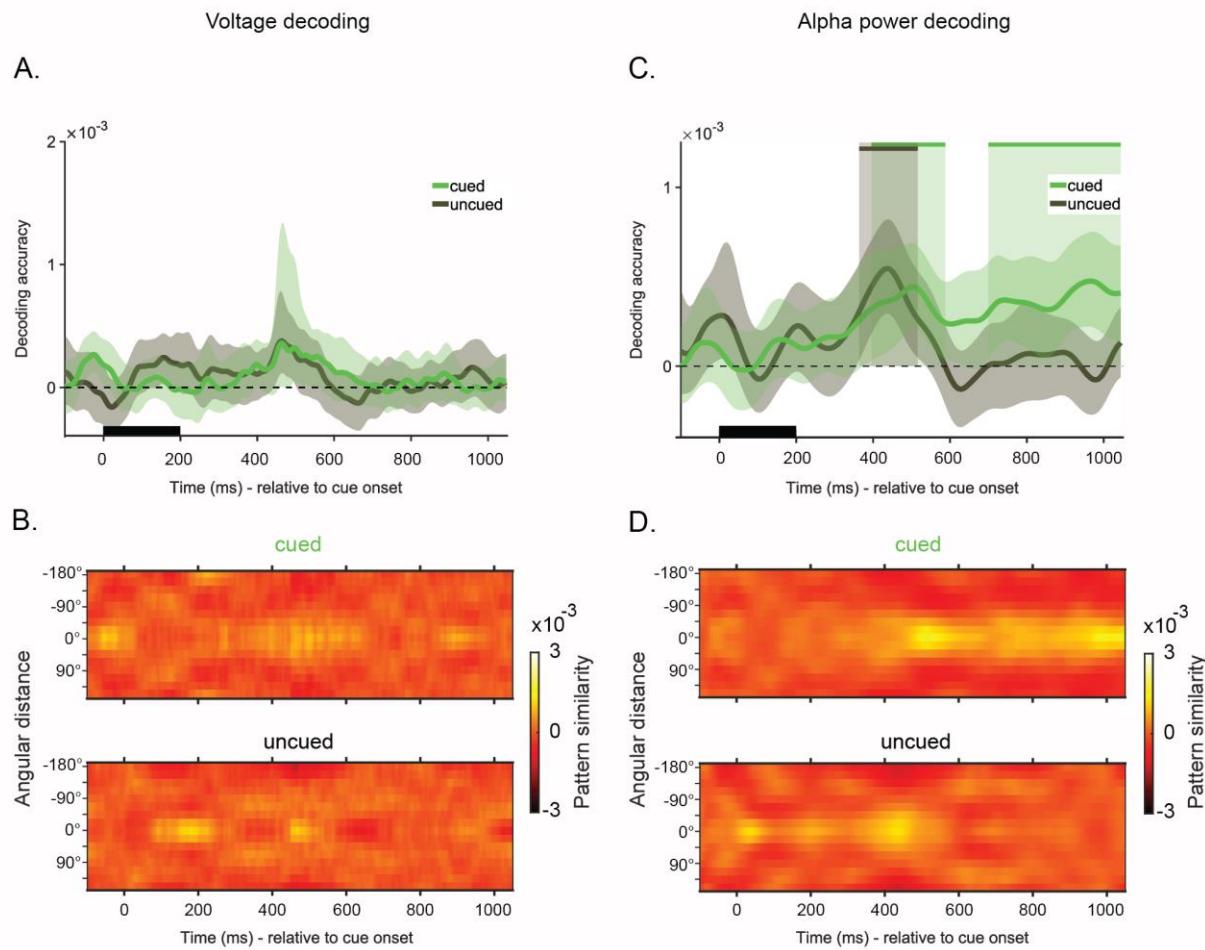
**Figure 4.** Time-course decoding of the cued (green) and the uncued (black) items, relative to the onset of the cue from the raw voltages **(A-B)**, and from alpha power **(C-D)**. **(A-C)** The black rectangular bar marks the presentation of the cue. Solid lines show the mean decoding accuracy (A.U.) over all trials and participants as a function of time. The shaded area around the mean marks the 95 % CI. Solid bars at the top and the shaded zones indicate statistically significant decoding periods ($p < 0.05$, one-sided). **(B-D)** Pattern similarity matrices for cued and uncued items show reverse-signed, mean-centred Mahalanobis distances between the target item and all other possible memory items, averaged over trials as a function of time.

18

At impulse presentation, voltage decoding revealed both the cued (Figure 5A, green, 28 ms to 514 ms, $p_{cued} < 0.001$, two-tailed, corrected), and the uncued item (Figure 5A, black, 164 ms to 236 ms, $p_{uncued} = 0.036$, two-tailed, corrected). The difference between the states of these two items was reflected by a significant difference in decoding accuracy (118 ms to 356 ms, $p_{cued - uncued} = 0.005$, one-tailed, corrected). Pattern similarity reflected this difference also, but was qualitatively similar for both items (Figure 5B). These results confirmed the outcomes of the analysis of the time window of interest reported above. The cued item was also decodable from alpha power (Figure 5C, green, 68 ms to 550 ms, $p < 0.001$, two-tailed, corrected), but the uncued item was not (difference; 92 ms to 548 ms, $p_{cued - uncued} < 0.001$, one-tailed, corrected). The pattern similarity matrix of the cued item showed high similarity near the tested values, with a steep drop-off further away (Figure 5D). These results suggest that while alpha power may reflect the sustained maintenance of, or the attention allocated to, the task-relevant memory item, the task-irrelevant item only emerged from the activity evoked by the impulse. Given that the uncued item also biased the eventual behavioural response, the present findings provide evidence that the impulse signal can reveal memory items retained in different activity states (Stokes, 2015; Stokes et al., 2020).

19
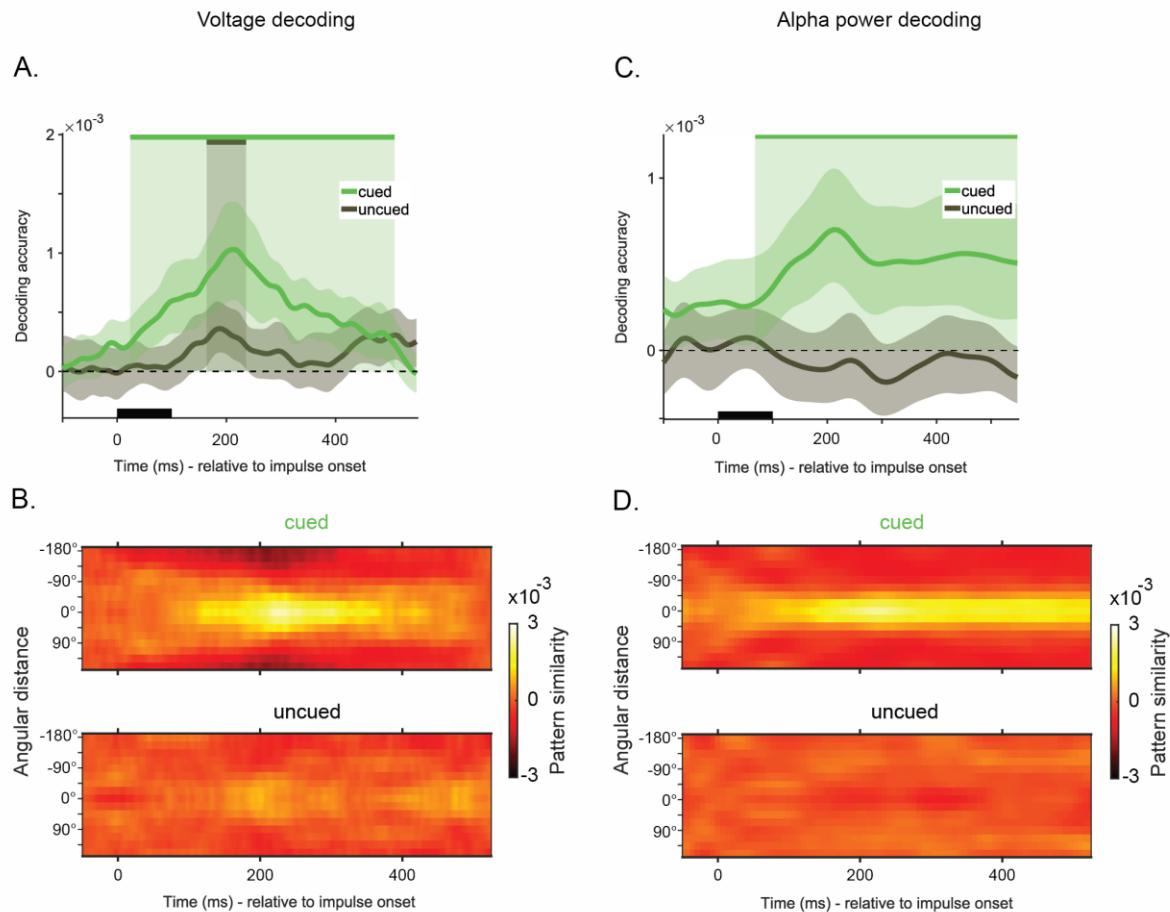
Figure 5. Time-course decoding of the cued (green) and the uncued (black) items, relative to impulse onset from the raw voltages **(A-B)**, and from alpha power **(C-D)**. **(A-C)** The black rectangular bar marks the presentation of the impulse. Solid lines show the mean decoding accuracy (A.U.) over all trials and participants as a function of time. The shaded area around the mean marks the 95 % CI. Solid bars at the top and the shaded zones indicate statistically significant decoding periods ($p < 0.05$, one-sided). **(B-D)** Pattern similarity matrices for cued and uncued items show reverse-signed, mean-centred Mahalanobis distances between the target item and all other possible memory items, averaged over trials as a function of time.

RSA of Colour Representations

20

In the RSA, two models were tested to explain pairwise differences between colour bins. The first model was based on a uniform colour space that reflected the angular differences between 16 colour bins on the colour wheel (Figure 6A, left). The second model was based on the behavioural reports (Figure 6A, right), in which errors were found to be more frequent around three primary colours (Figure 2B). In the model, these three colours were thus used to explain the variation in the data. The average RDMs presented in Figure 6B reflect the pairwise differences between the 16 colour bins within the time window of 100 to 400 ms, relative to the onset of the memory items (left), and relative to impulse onset (right).

Both models fit the data during stimulus encoding (Figure 6C, left, $p_{\text{Uniform colours}} < 0.001$, one-tailed; $p_{\text{Primary colours}} < 0.001$, one-tailed). Thus, the results provided evidence for a parametric relationship between colour representations during sensory encoding, as well as a degree of primary-colour categorization. The uniform colour space model also fit the neural response associated with the cued item that was evoked by the impulse (Figure 6C, right, $p_{\text{Uniform colours}} < 0.001$, one-tailed), in line with earlier reports (Hajonides et al, 2021). Conversely, although there was a trend, the primary colours model failed to reach statistical significance (Figure 6C, right, $p_{\text{Primary colours}} = 0.054$, one-tailed).
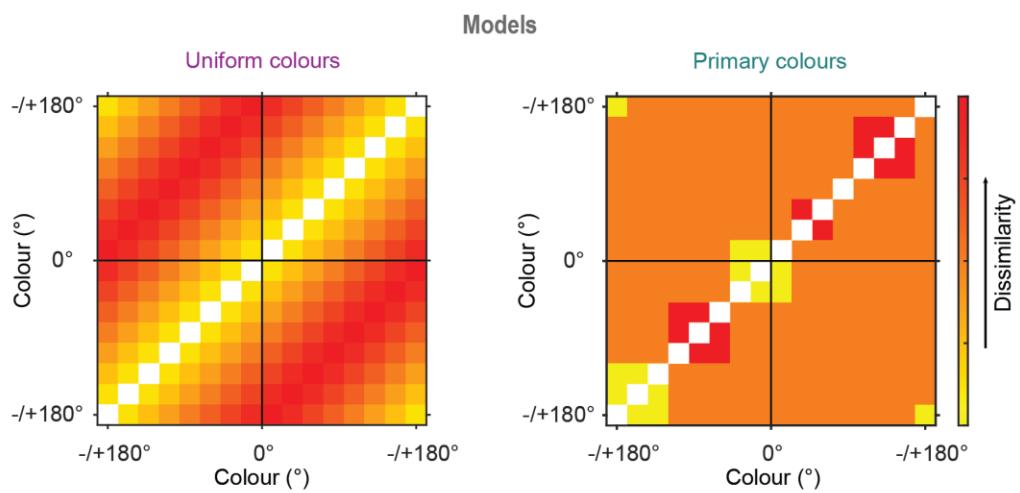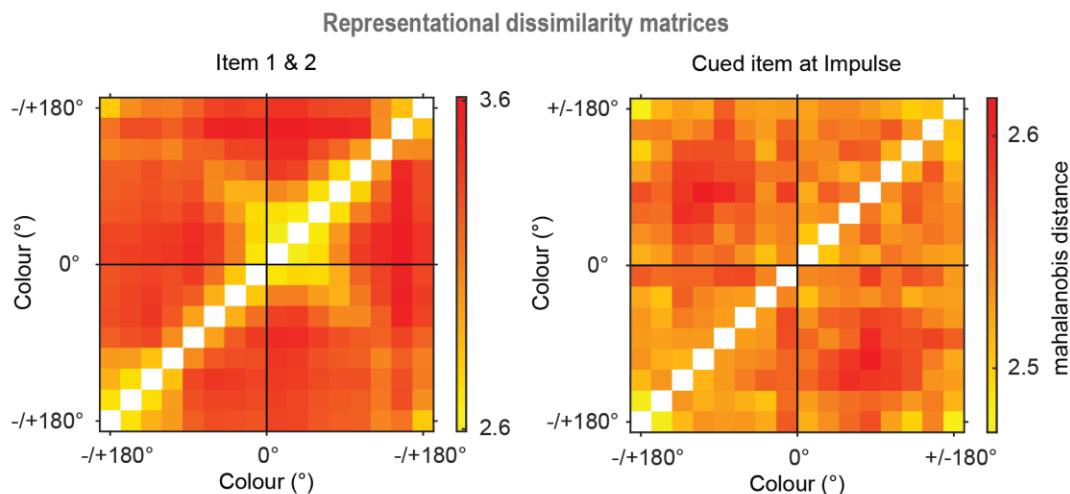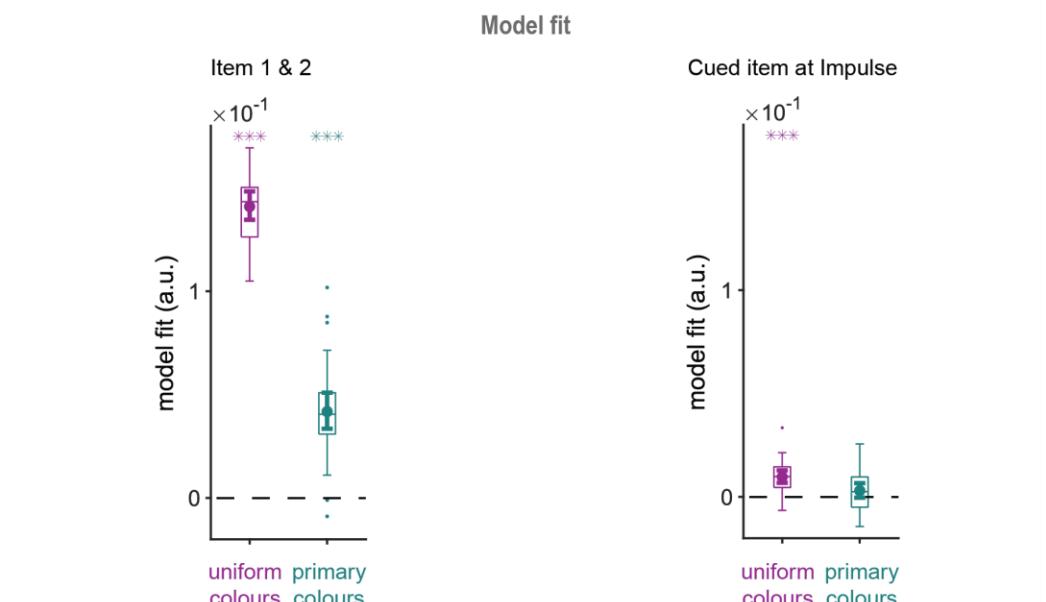
Figure 6. Representational similarity analysis of the average of item 1 and 2 (left) and the cued item at impulse (right). **(A)** Model for a uniform, continuous colour space for 16 colour bins (left), and model for a discrete colour space based on primary colours extracted from behavioural errors (right). **(B)** The representational dissimilarity matrices (RDM) for item presentation (left) and impulse (right) epochs. **(C)** Boxplots showing the mean fit (beta values) following the linear regression of individual RDMs on the models for each participant. The dot in the centre marks the mean standardized slope (beta), and the whiskers on both ends of the mean indicate the 95% CI. The boxes border the 25th and 75th percentiles, with the whiskers around the box stretching to 1.5 interquartile range. Asterisks indicate beta values that were significantly above zero (* , $p < 0.05$; ** , $p < 0.01$; *** , $p < 0.001$).

Discussion

We aimed to decode colours maintained in working memory by means of visual impulse perturbation. In a pre-registered experiment, we tested the maintenance of a retro-actively cued target item, and of the item that was not cued, in a delayed match-to-sample task. To further avoid possible spatial confounds that might have affected previous impulse perturbation experiments, we presented our two memory items serially in the centre of the screen, and randomized the appearance of the colour circle shown at the probe on each trial. We were able to decode stimulus identity from the recorded EEG signal during perceptual encoding, replicating earlier studies (Bocincova & Johnson, 2019; Brouwer & Heeger, 2009; Hajonides et al., 2021; Hermann et al., 2022; Persichetti et al., 2015; Rosenthal et al., 2021; Sandhaeger et al., 2019). Importantly, we could also decode stimulus identity post-impulse. This result extended earlier studies that decoded the bottom-up activity induced by the sensory processing of a visual impulse signal to reveal representations of orientation gratings in working memory that were otherwise not traceable from raw EEG (e.g., Wolff et. al., 2015; 2017; 2020). The present result was the first demonstration of impulse-driven decoding of non-spatial features, specifically colour memories, thereby generalizing the findings across feature dimensions.

23

In contrast to previous studies (Kuo et al., 2012; LaRocque et al., 2013; Lewis-Peacock et al., 2012; Wolff et al., 2017; 2020a; 2020b), the impulse effect was not restricted to the task-relevant cued item. This suggests that the task-irrelevant uncued item was also maintained in memory, albeit to a lesser degree. One trivial reason for this might be that participants simply selected the wrong item on some trials, but there is reason to doubt this explanation. First, in alpha power decoding (discussed further below), the signal associated with the uncued item behaved completely different from the cued item, both during maintenance and following the impulse. Second, behavioural response errors on the cued item were biased by the uncued item (cf. Bae& Luck, 2017; Golomb, 2015; Huang & Sekuler, 2010), but they were biased away from it, rather than towards it—the latter would be expected when the uncued item was erroneously reported.

The emergence of the uncued item after impulse onset casts doubt on the idea that information in working memory that is no longer needed is actively purged. Memory might alternatively 'let go of' task-irrelevant items, for instance by no longer periodically refreshing them, such that their representations fade relatively quickly, but might nevertheless still linger for some time. This idea is compatible also with a recent model of working memory based on calcium-mediated short-term synaptic plasticity (Pals et al., 2020). The absence of task-irrelevant items in neural measures obtained in previous studies might be a consequence of the inherent weakness of their representations, which makes them harder to detect than task-relevant items in the first place. The high number of trials in the current study may have helped to overcome this difficulty. However, from the present data it cannot be excluded that the presence of the uncued item during memory maintenance might be related specifically to the metathetic nature of colour. This awaits further experimentation.

The decoding analyses showed that alpha band decoding and dynamic voltage decoding clearly reflected different states of the memoranda. After cue onset, item selection following the processing of the cue was only observable in alpha power. Both the task-relevant and irrelevant colours could initially be decoded, followed by a sustained signal for the cued item only. With regard to the cued item, the presentation of the cue would be hypothesised to require an active process of prioritizing one item over the other, or potentially even removing the uncued item from memory storage (but see the argument above).

24

Even theories of activity-silent working memory have so far proposed connectivity-based, synaptic storage only as a mechanism of maintenance, and updating of synaptic weights would still require neuronal firing (Kozachkov et al., 2022; Pals et al., 2020; Stokes et al., 2015; Trubutschek et al., 2019; Wolff et al., 2021). Based on the analyses of our current dataset it appears that this attentional selection happens uniquely in the alpha band, which subsequently then also holds the prioritized item in a sustained manner for the rest of the trial. With regard to the uncued item, the absence of a sustained signal both in the raw EEG and in the alpha band suggests that it was truly activity-quiescent, if not altogether silent.

Once the impulse was presented, the voltage decoding revealed both the cued and uncued item, contrary to the alpha band decoding. Additionally, the reactivation of the signal in the voltage decoding was time-limited, and returned to zero well before the presentation of the probe. This provides further evidence that the alpha band and voltage decoding track functionally different states of working memory. We speculate that the alpha band may serve to keep memoranda in an elevated state, ready for direct access when necessary (Oberauer, 2001; 2002). It seems likely that attention mediates this (Bae & Luck, 2018). Following the cue, the uncued item was apparently removed from this elevated state, but it was not altogether lost, as dynamic voltage decoding still revealed it from the EEG response to the impulse. Although the functional states of items in memory need not necessarily map directly onto corresponding neural states (Muhle-Karbe et al., 2021; Stokes et al., 2020), the task-irrelevant uncued item was clearly in a different, more silent, neural state than the task-relevant cued item in our data. The fact that we can trace these differences and differentiate them provides further support for the utility of the impulse driven decoding technique.

Finally, we found that the neural representations of colours were parametrically arranged during encoding (Brouwer & Heeger, 2009; Hajonides et al., 2021). The signal evoked by the impulse also showed a parametric arrangement of colours during the delay period. Nevertheless, behavioural responses reflected a bias in reports, as the errors were grouped around three primary colours. This observation was in line with earlier reports that not all colours are equally memorable (Bae et al, 2014). A discrete colour model derived from the behavioural output was indeed also supported during the encoding of the memory items (i.e.,

25

directly after their onset), suggesting that participants formed a categorical representation for colours even during perception or shortly thereafter. Crucially, the evidence for this discrete model was no longer reliable during memory maintenance (i.e., after the impulse), while the continuous model was still well supported. It may appear paradoxical that the categorical representation observed during encoding, and in the eventual response, was not clearly represented during memory maintenance. One possible explanation for this could be the differential impact of the activity induced by the sensory processing of the impulse signal on the different networks that might retain continuous colour representations and discrete categorical representations. One might suppose that more discrete colour groupings could be represented by a semantic network (Rosenthal, et al., 2021), whereas more low-level hue differences could be retained in earlier visual areas (Brouwer & Heeger, 2009). The latter might be more accessible to impulse perturbation.

Conclusion

The results of our study highlighted both commonalities and potential differences between spatial and non-spatial (colour) items maintained in working memory. First, we observed that the visual impulse response contained information not only about the task-relevant, cued item, but also the task-irrelevant, uncued item, contrary to previous studies of orientation gratings. This finding casts doubt on the idea that uncued items are actively purged from memory. Second, we found evidence for a sustained signal corresponding to the cued, but not the uncued, item during the delay period. This suggests that the alpha band may uniquely trace the attended item that is in the focus of attention. Furthering the debate on the relationship between the functional role of an item and its activity state (Barbosa et al., 2021; Stokes et al., 2020; Wolff et al., 2021), the current results provided evidence for the utility of the impulse technique by revealing silent memories that were not only hidden in ongoing EEG, but also untraceable by alpha power.

Acknowledgements

27

References

Bae, G. Y., & Luck, S. J. (2017). Interactions between visual working memory representations. *Attention Perception, & Psychophysics, 79*(8), 2376–2395. https://doi.org/10.3758/s13414-017-1404-8

Bae, G. Y., & Luck, S. J. (2018). Dissociable decoding of spatial attention and working memory from EEG oscillations and sustained potentials. *Journal of Neuroscience, 38*(2), 409–422. https://doi.org/10.1523/JNEUROSCI.2860-17.2017

Bae, G., Olkkonen, M., Allred, S. R., Wilson, C., & I. Flombaum, J. I. (2014). Stimulus-specific variability in color working memory with delayed estimation. *Journal of Vision*, *14*(4), 7. https://doi.org/10.1167/14.4.7

Barak, O., & Tsodyks, M. (2014). Working models of working memory. *Current Opinion in Neurobiology*, *25*, 20–24. https://doi.org/10.1016/j.conb.2013.10.008

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. https://doi.org/10.1371/journal.pbio.3001436

Bocincova, A., & Johnson, J. S. (2019). The time course of encoding and maintenance of task-relevant versus irrelevant object features in working memory. *Cortex, 111*, 196–209. https://doi.org/10.1016/j.cortex.2018.10.013

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436. https://doi.org/10.1163/156856897X00357

Brouwer, G. J., & Heeger, D. J. (2009). Decoding and reconstructing color from responses in human visual cortex. *Journal of Neuroscience, 29*(44), 13992–14003. https://doi.org/10.1523/JNEUROSCI.3577-09.2009

Buonomano, D. V., & Maass, W. (2009). State-dependent computations: Spatiotemporal processing in cortical networks. *Nature Reviews Neuroscience*, *10*(2), 113–125. https://doi.org/10.1038/nrn2558

28

Constantinidis, C., Funahashi, S., Lee, D., Murray, J. D., Qi, X.-L., Wang, M., & Arnsten, A. F. T. (2018). Persistent spiking activity underlies working memory. *Journal of Neuroscience*, *38*(32), 7020–7028. http://doi.org/10.1523/JNEUROSCI.2486-17.2018

Cowan, N. (2017). The many faces of working memory and short-term storage. *Psychonomic Bulletin & Review*, *24*(4), 1158–1170. https://doi.org/10.3758/s13423-016-1191-6

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*(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

De Maesschalck, R., Jouan-Rimbaud, D., & Massart, D. L. (2000). The Mahalanobis distance. *Chemometrics and Intelligent Laboratory Systems*, *50*(1), 1–18. https://doi.org/10.1016/S0169-7439(99)00047-7

Emery, K. J., & Webster, M.A. (2019). Individual differences and their implications for color perception. *Current Opinion in Behavioral Sciences, 30,* 28-33. https://doi.org/10.1016/j.cobeha.2019.05.002

Fan, Y., Han, Q., Guo, S., & Luo, H. (2021). Distinct neural representations of content and ordinal structure in auditory sequence memory. *Journal of Neuroscience*, *41*(29), 6290-6303. https://doi.org/10.1523/JNEUROSCI.0320-21.2021

Fan, Y., & Luo, H. (2022). Reactivating ordinal position information from auditory sequence memory in human brains. *Cerebral Cortex*, bhac471. https://doi.org/10.1093/cercor/bhac471

Fornaciai, M., & Park, J. (2020). Neural dynamics of serial dependence in numerosity perception. *Journal of Cognitive Neuroscience, 32*(1), 141–154. https://doi.org/10.1162/jocn_a_01474

Goldman-Rakic P. S. (1995). Cellular basis of working memory. *Neuron*, *14*(3), 477–485. https://doi.org/10.1016/0896-6273(95)90304-6

Golomb, J. D. (2015). Divided spatial attention and feature mixing errors. *Attention, Perception, & Psychophysics, 77*(8), 2562–2569. https://doi.org/10.3758/s13414-015-0951-0.

Hajonides, J. E., Nobre, K., van Ede, F., & Stokes, M. G. (2021). Decoding visual color from scalp electroencephalography measurements. *NeuroImage, 237*, 118030. https://doi.org/10.1016/j.neuroimage.2021.118030

Hermann, K. L., Singh, S. R., Rosenthal, I. A., Pantazis, D., & Conway, B. R. (2022). Temporal dynamics of the neural representation of hue and luminance polarity. *Nature Communications, 13*, 661. https://doi.org/10.1038/s41467-022-28249-0

Huang, J., & Sekuler, R. (2010). Distortions in recall from visual memory: Two classes of attractors at work. *Journal of Vision, 10*(2), 24. https://doi.org/10.1167/10.2.24

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in Psychtoolbox-3. *Perception, 36*(14), 1-16.

Kojima, S., & Goldman-Rakic, P. S. (1982). Delay-related activity of prefrontal neurons in rhesus monkeys performing delayed response. *Brain Research*, *248*(1), 43–50. https://doi.org/10.1016/0006-8993(82)91145-3

Kozachkov, L., Tauber, J., Lundqvist, M., Brincat, S. L., Slotine, J-J., & Miller, E. K. (2022). Robust and brain-like working memory through short-term synaptic plasticity. *PLoS Computational Biology*, *18*(12), e1010776. https://doi.org/10.1371/journal.pcbi.1010776

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis - connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*, 4. https://doi.org/10.3389/neuro.06.004.2008

Kuo, B.-C., Stokes, M. G., & Nobre, A. C. (2012). Attention modulates maintenance of representations in visual short-term memory. *Journal of Cognitive Neuroscience*, 24(1), 51–60. https://doi.org/10.1162/jocn_a_00087

LaRocque, J. J., Lewis-Peacock, J. A., Drysdale, A. T., Oberauer, K., & Postle, B. R. (2013). Decoding attended information in short-term memory: An EEG study. *Journal of Cognitive Neuroscience*, *25*(1), 127–142. https://doi.org/10.1162/jocn_a_00305

Ledoit, O., & Wolf, M. (2004). Honey, I shrunk the sample covariance matrix. *Journal of Portfolio Management*, *30*,110–119. https://doi.org/10.3905/jpm.2004.110

Lewis-Peacock, J. A., Drysdale, A. T., Oberauer, K., & Postle, B. R. (2012). Neural evidence for a distinction between short-term memory and the focus of attention. *Journal of Cognitive Neuroscience*, *24*(1), 61–79. https://doi.org/10.1162/jocn_a_00140

Lundqvist, M., Herman, P., & Miller, E. K. (2018). Working memory: Delay activity, yes! Persistent activity? Maybe not. *Journal of Neuroscience, 38*(32), 7013– 7019. http://doi.org/10.1523/JNEUROSCI.2485-17.2018

Miller, E. K., Lundqvist, M., & Bastos, A. M. (2018). Working memory 2.0. *Neuron, 100*(2), 463–475. https://doi.org/10.1016/j.neuron.2018.09.023

Mongillo, G., Barak, O., & Tsodyks, M. (2008). Synaptic theory of working memory. *Science*, *319*(5869), 1543–1546. https://doi.org/10.1126/science.1150769

Mostert, P., Albers, A. M., Brinkman, L., Todorova, L., Kok, P., & de Lange, F. P. (2018). Eye movement-related confounds in neural decoding of visual working memory representations, *eNeuro*, *5*. https://doi.org/10.1523/ENEURO.0401-17.2018

Muhle-Karbe, P. S., Myers, N. E., & Stokes, M. G. (2021). A hierarchy of functional states in working memory. *Journal of Neuroscience, 41*(20), 4461–4475. https://doi.org/10.1523/JNEUROSCI.3104-20.2021

Nishimoto, S., Huth, A. G., Bilenko, N. Y., & Gallant, J. L. (2017). Eye movement-invariant representations in the human visual system. *Journal of Vision, 17*(1), 11. https://doi.org/10.1167/17.1.11

Oberauer, K. (2001). Removing irrelevant information from working memory: A cognitive aging study with the modified Sternberg task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(4), 948–957. https://doi.org/10.1037/0278-7393.27.4.948

Oberauer, K. (2002). Access to information in working memory: Exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(3), 411–421. https://doi.org/10.1037/0278-7393.28.3.411

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, *2011*, 156869. https://doi.org/10.1155/2011/156869

Pals, M., Stewart, T. C., Akyürek, E. G., & Borst, J. P. (2020). A functional spiking-neuron model of activity-silent working memory in humans based on calcium-mediated short-term synaptic plasticity. *PLoS Computational Biology, 16*, e1007936. https://doi.org/10.1371/journal.pcbi.1007936

Persichetti, A. S., Thompson-Schill, S. L., Butt, O. H., Brainard, D. H., & Aguirre, G. K. (2015). Functional magnetic resonance imaging adaptation reveals a noncategorical representation of hue in early visual cortex. *Journal of Vision, 15*(6), 18. https://doi.org/10.1167/15.6.18

Quax, S. C., Dijkstra, N., van Staveren, M. J., Bosch, S. E., & van Gerven, M. A. (2019). Eye movements explain decodability during perception and cued attention in MEG. *NeuroImage*, *195*, 444–453. https://doi.org/10.1016/j.neuroimage.2019.03.069

Rosenthal, I. A., Singh, S. R., Hermann, K. L., Pantazis, D., & Conway, B. R. (2021). Color space geometry uncovered with magnetoencephalography. *Current Biology*, *31*, 515-526.

Sandhaeger, F., von Nicolai, C., Miller, E. K., & Siegel, M. (2019). Monkey EEG links neuronal color and motion information across species and scales. *eLife*, *8*, e45645. https://doi.org/10.7554/eLife.45645

Shafi, M., Zhou, Y., Quintana, J., Chow, C., Fuster, J., & Bodner, M. (2007). Variability in neuronal activity in primate cortex during working memory tasks. *Neuroscience*, *146*(3), 1082–1108. https://doi.org/10.1016/j.neuroscience.2006.12.072

Stokes, M. G., Wolff, M. J., & Spaak, E. (2015). Decoding rich spatial information with high temporal resolution. *Trends in Cognitive Sciences*, *19*(11), 636–638. https://doi.org/10.1016/j.tics.2015.08.016

Stokes, M. G., Muhle-Karbe, P. S., & Myers, N. E. (2020). Theoretical distinction between functional states in working memory and their corresponding neural states. *Visual Cognition, 28*, 420–432. https://doi.org/10.1080/13506285.2020.1825141

Suchow, J. W., Brady, T. F., Fougnie, D., & Alvarez, G. A. (2013). Modeling visual working memory with the MemToolbox. *Journal of Vision*, *13*(10), 9. https://doi.org/10.1167/13.10.9.

Thielen, J., van Lier, R. & van Gerven, M. (2018). No evidence for confounding orientation-dependent fixational eye movements under baseline conditions, *Scientific Reports, 8*, 11644. https://doi.org/10.1038/s41598-018-30221-2

Thielen, J., Bosch, S.E., van Leeuwen, T.M., van Gerven, M., & van Lier, R. (2019). Evidence for confounding eye movements under attempted fixation and active viewing in cognitive neuroscience, *Scientific Reports*, *9*, 17456. https://doi.org/10.1038/s41598-019-54018-z

Trübutschek, D., Marti, S., Ojeda, A., King, J-R., Mi, Y., Tsodyks, M., & Dehaene, S. (2017). A theory of working memory without consciousness or sustained activity. *eLife*, *6*, e23871. https://doi.org/10.7554/eLife.23871

Watanabe, K., & Funahashi, S. (2014). Neural mechanisms of dual-task interference and cognitive capacity limitation in the prefrontal cortex. *Nature Neuroscience*, *17*(4), 601–611. https://doi.org/10.1038/nn.3667

Wolff M. J., Ding, J., Myers, N. E., & Stokes, M. G. (2015). Revealing hidden states in visual working memory using electroencephalography. *Frontiers in Systems Neuroscience, 9*, 123. https://doi.org/10.3389/fnsys.2015.00123

Wolff, M. J., Jochim, J., Akyürek, E. G., & Stokes, M. G. (2017). Dynamic hidden states underlying working-memory-guided behavior. *Nature Neuroscience*, *20*(6), 864–871. https://doi.org/10.1038/nn.4546

Wolff, M. J., Kandemir, G., Stokes, M. G., & Akyürek, E. G. (2020a). Unimodal and bimodal access to sensory working memories by auditory and visual impulses. *Journal of Neuroscience*, *40*(3), 671–681. https://doi.org/10.1523/JNEUROSCI.1194-19.2019

Wolff, M. J., Jochim, J., Akyürek, E. G., Buschman, T. J., & Stokes, M. G. (2020b). Drifting codes within a stable coding scheme for working memory, *PLoS Biology*, *18*(3), e3000625. https://doi.org/10.1371/journal.pbio.3000625

Wolff, M. J., Akyürek, E. G., & Stokes, M. G. (2021). What is the functional role of delay-related alpha oscillations during working memory? *PsyArXiv*. https://doi.org/10.31234/osf.io/z6y5b

Worden, M.S., Foxe, J.J., Wang, N., & Simpson, G.V. (2000). Anticipatory biasing of visuospatial attention indexed by retinotopically specific alpha-band electroencephalography increases over occipital cortex. *Journal of Neuroscience, 20*, RC63. https://doi.org/10.1523/JNEUROSCI.20-06-j0002.2000

Supplementary Information

To ensure that the decoding of the memory items was not confounded by (attentionally driven) eye movements, we decoded the bipolar voltage measures taken from the eye electrodes. The decoding approach was identical to that of the time course analysis of the EEG. The statistical analyses revealed no statistically significant clusters for which decoding accuracy was above zero.
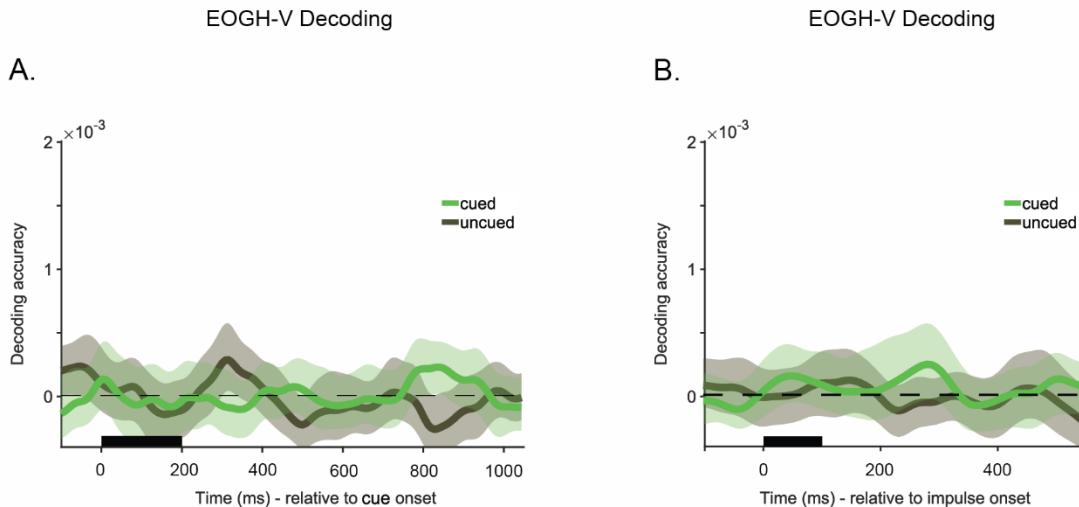

Figure S1. Time-course decoding of the cued (green) and the uncued (black) items, relative to cue onset (A) and impulse onset (B) from the voltage measures from the bipolar eye electrodes. The black rectangular bar marks the presentation of the cue (A) and the impulse (B). Solid lines show the mean decoding accuracy over all trials and participants as a function of time. The shaded area around the mean marks the 95 % CI.