

Cognition and Behavior

# Eye Movement-Related Confounds in Neural Decoding of Visual Working Memory Representations

Pim Mostert,<sup>1</sup>  Anke Marit Albers,<sup>1,2</sup>  Loek Brinkman,<sup>1,3</sup>  Larisa Todorova,<sup>1,4</sup>  Peter Kok,<sup>5,6</sup> and  Floris P. de Lange<sup>1</sup>

DOI:<http://dx.doi.org/10.1523/ENEURO.0401-17.2018>

<sup>1</sup>Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen 6500 HB, The Netherlands, <sup>2</sup>Abteilung Allgemeine Psychologie, Justus-Liebig-Universität, Giessen 35394, Germany, <sup>3</sup>Department of Psychology, Utrecht University, Utrecht 3584 CS, The Netherlands, <sup>4</sup>Center for Mind/Brain Sciences, University of Trento, Rovereto TN 38068, Italy, <sup>5</sup>Princeton Neuroscience Institute, Princeton University, Princeton, NJ 08544, and <sup>6</sup>Department of Psychology, Yale University, New Haven, CT 06511

## Abstract

A relatively new analysis technique, known as neural decoding or multivariate pattern analysis (MVPA), has become increasingly popular for cognitive neuroimaging studies over recent years. These techniques promise to uncover the representational contents of neural signals, as well as the underlying code and the dynamic profile thereof. A field in which these techniques have led to novel insights in particular is that of visual working memory (VWM). In the present study, we subjected human volunteers to a combined VWM/imagery task while recording their neural signals using magnetoencephalography (MEG). We applied multivariate decoding analyses to uncover the temporal profile underlying the neural representations of the memorized item. Analysis of gaze position however revealed that our results were contaminated by systematic eye movements, suggesting that the MEG decoding results from our originally planned analyses were confounded. In addition to the eye movement analyses, we also present the original analyses to highlight how these might have readily led to invalid conclusions. Finally, we demonstrate a potential remedy, whereby we train the decoders on a functional localizer that was specifically designed to target bottom-up sensory signals and as such avoids eye movements. We conclude by arguing for more awareness of the potentially pervasive and ubiquitous effects of eye movement-related confounds.

**Key words:** eye movements; magnetoencephalography; multivariate decoding; visual working memory

## Significance Statement

Neural decoding is an important and relatively novel technique that has opened up new avenues for cognitive neuroscience research. However, with its promises also come potential caveats. In this study we show that neural decoding may be susceptible to confounds induced by small task- and stimulus-specific eye movements in the context of a visual working memory (VWM) task. Such eye movements during working memory tasks have been reported before and may in fact be a common phenomenon. Given the widespread use of neural decoding and the potentially contaminating effects of eye movements, we therefore believe that our results are of significant relevance for the field.

## Introduction

Neural decoding, or multivariate pattern analysis (MVPA), is a popular analysis technique that has obtained considerable momentum in the field of cognitive neuroimaging (Haxby et al., 2014; Grootswagers et al., 2017). It refers to uncovering a factor of interest, for instance stimulus identity, from multivariate patterns in neural signals such as those measured by magnetoencephalography (MEG) or functional magnetic resonance imaging (fMRI). Decoding allows one to probe the representational contents of a neural signal, rather than overall activity levels, with superior sensitivity. However, this sensitivity may require extra vigilance at the end of the user, because these analyses may also be particularly sensitive to potentially confounding factors. Here we demonstrate such an example, specifically in the context of visual working memory (VWM), where a decoding analysis is contaminated by stimulus-specific eye movements. Given the widespread use of these techniques and its pivotal contributions to contemporary VWM theories, we argue that appreciation of these potential caveats is important.

VWM is the ability to retain and use visual information about the world for a short period of time, even when the original external source of that information is no longer available. Neural decoding has been frequently applied in the study of VWM to elucidate where, when and how a memorandum is encoded in the brain. This was first demonstrated by Harrison and Tong (2009) and Serences et al. (2009), who were able to decode the orientation of a memorized grating from visual cortex. Further VWM decoding studies extended Harrison and Tong (2009)'s paradigm in varying ways to study, among others, mental imagery, mental transformations, and spatial working memory (Albers et al., 2013; Christophel et al., 2015, 2017; Foster et al., 2016; Gayet et al., 2017). The paradigm has also been ported to electrophysiological studies using MEG or electroencephalography to capitalize on the high temporal resolution offered by those methods (Wolff et al., 2015, 2017; Foster et al., 2016; King et al., 2016). These results have led to important new theories, among others the idea that high-fidelity VWM representations are

stored in early sensory cortex (Albers et al., 2013; Sreenivasan et al., 2014), the activity-silent coding hypothesis (Stokes, 2015; Wolff et al., 2015, 2017; Rose et al., 2016; Rademaker and Serences, 2017) and the dynamic coding framework (Stokes et al., 2013, 2015; King et al., 2016; Spaak et al., 2017).

In the current study, human volunteers performed a combined VWM/imagery task, while we traced the representational contents of their neural activity as measured by MEG. The experiment was designed to elucidate the temporal profile of the memorized item's neural representation. However, control analyses revealed that our data were severely contaminated by small eye movements. In this article, we first describe the eye movement analysis to show how the identity of the memorized item could be decoded from gaze position. Next, we present the naive results as they would have been, had we not been aware of the confound. This highlights how these could easily have been mistaken to provide genuine insight into the neural mechanisms underlying VWM. Finally, we present a potential solution by training the decoders on separate functional localizer blocks, which allowed us to extract the sensory-specific neural patterns, thereby effectively bypassing the eye-movements confounds.

## Materials and Methods

### Subjects

Thirty-six human volunteers were recruited from the local institute's subject pool to participate in a behavioral screening session. Of these, 24 (13 male; mean age: 26.8 years, range: 18–60) were selected to participate in the MEG experiment (see below, Experimental design and procedure). Of these 24 selected subjects, three were excluded from MEG analysis due to poor data quality and another four were excluded from the analyses regarding eye movements, because the eye-tracker failed to track the eye reliably in those subjects. The experiment was approved by the local ethics committee and conducted according to the guidelines set out by the committee. All participants provided written informed consent and received either monetary compensation or course credits.

### Stimuli

Stimulation was visual and consisted of sinusoidal gratings with a spatial frequency of 1 cycle/°, 80% contrast and one random phase per experimental block. The gratings were masked at an outer radius of 7.5° and an inner aperture radius of 0.7° and presented on a gray background (luminance: 186 cd/m<sup>2</sup>). Stimuli were generated and presented using MATLAB with the Psychtoolbox extension (Kleiner et al., 2007).

### Experimental design and procedure

The main task was to vividly imagine and remember an oriented grating and, in some conditions, mentally rotate this grating over a certain angle. Each trial began with a dual cue that indicated both the amount (presented above fixation) and the direction (> for clockwise and < for counterclockwise, presented below fixation) of mental rotation (MR) that was to be performed in that trial (Fig. 1).

Received November 21, 2017; accepted June 12, 2018; First published July 18, 2018.

The authors declare no competing financial interests.

Author contributions: A.M.A., L.B., L.T., and F.P.d.L. designed research; A.M.A. and L.T. performed research; P.M. analyzed data; P.M., A.M.A., L.B., P.K., and F.P.d.L. wrote the paper.

This work was supported by The Netherlands Organisation for Scientific Research and the European Research Council [NWO Research Talent Grant 406-13-001 (to P.M.); NWO Brain and Cognition Grant 433-09-248 (to A.M.A. and L.B.); NWO Rubicon Grant 446-15-004 (to P.K.); and NWO Vidi Grant 452-13-016, ERC Starting Grant 678286, and NWO Brain and Cognition Grant 433-09-248 (to F.P.d.L.)].

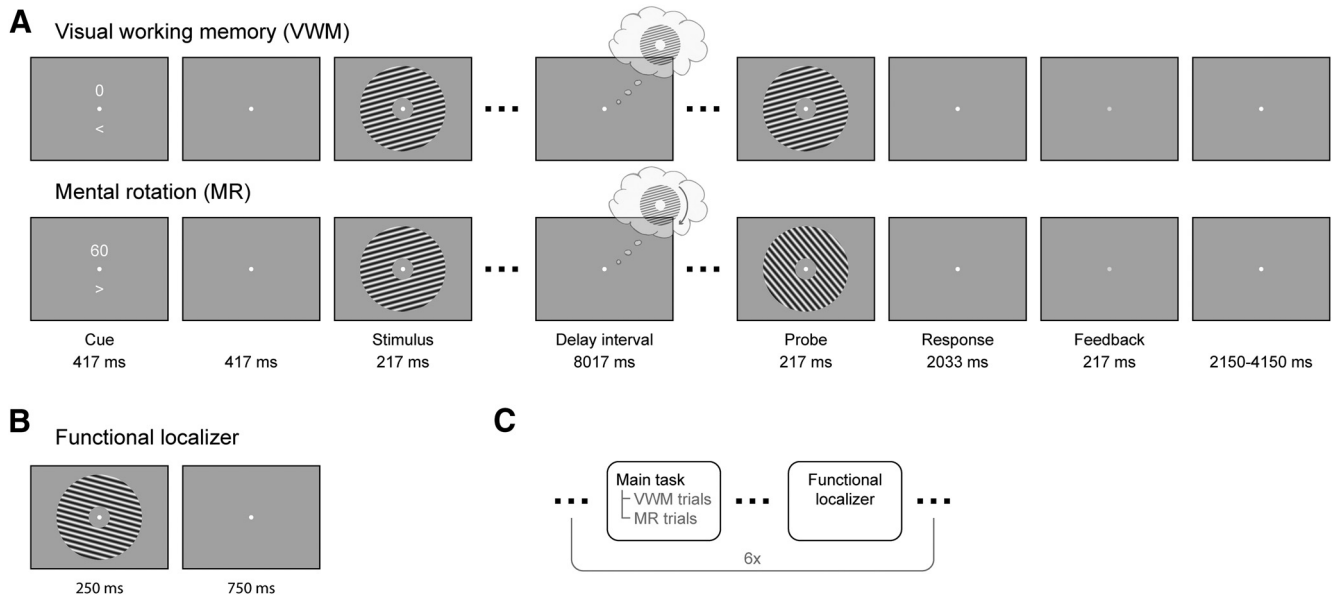
Acknowledgements: We thank Ivan Toni for useful suggestions during the design phase.

Correspondence should be addressed to Floris P. de Lange at the above address, E-mail: [floris.delange@donders.ru.nl](mailto:floris.delange@donders.ru.nl).

DOI:<http://dx.doi.org/10.1523/ENEURO.0401-17.2018>

Copyright © 2018 Mostert et al.

This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license, which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed.



**Figure 1.** Experimental paradigm. **A**, In the combined VWM/imagery blocks, subjects were instructed to vividly imagine a grating and to either keep that in mind (VWM condition) or rotate it mentally over a cued number of degrees (MR condition). **B**, In the functional localizer block, oriented gratings were continuously presented while the subject's attention was drawn to a task at fixation. **C**, The VWM/imagery and localizer blocks were performed in alternating order.

The amount could be either  $0^\circ$ ,  $60^\circ$ ,  $120^\circ$ , or  $180^\circ$ , in either clockwise or counterclockwise direction, where  $0^\circ$  corresponded to a VWM task. This condition will henceforth be referred to as the VWM condition, and the other three conditions, which corresponded to imagery, as the MR conditions. This cue lasted for 417 ms, after which a blank screen was shown for another 417 ms. A fixation dot (four pixels in diameter) was present throughout the entire trial, and throughout the entire block. After the blank, a grating was presented for 217 ms that could have either of three orientations:  $15^\circ$ ,  $75^\circ$ , or  $135^\circ$  (clockwise with respect to vertical). Next, a blank delay period of 8017 ms followed, during which subjects were required to keep the starting grating in mind and, in a subset of trials, mentally rotate it. The delay period was terminated by the presentation of a probe grating for 217 ms, whose orientation was slightly jittered (see below, Staircase procedure) with respect to the orientation that subjects were supposed to have in mind at that moment. Subjects then indicated with a button press whether the probe was oriented clockwise or counterclockwise relative to their internal image. The response period lasted until 2033 ms after probe, after which feedback was given. There were three trials per design cell (four arcs of rotation and two directions) per block, resulting in 24 trials per block. In addition, there were two catch trials per block, in which the probe grating was presented at an earlier moment in the delay interval to gauge ongoing rotation. All trials were presented in pseudorandomized order. The catch trials were excluded from further analysis, because subjects indicated to find them difficult and confusing. In general, each experiment consisted of six experimental blocks (although some subjects performed five, seven, or eight blocks), preceded by one or more practice blocks, resulting in a total of 144 experimental trials for most of the participants.

Interleaved with the VWM/imagery blocks, there were six functional localizer blocks (Fig. 1B,C). In these blocks, gratings of six different orientations ( $15^\circ$  to  $165^\circ$ , in steps of  $30^\circ$ ) were presented for 250 ms with an intertrial interval of 750 ms. Each block consisted of 120 trials, resulting in a total of 120 trials per orientation. The task was to press a button when a brief flicker of the fixation dot occurred. Such a flicker occurred between 8 and 12 times (randomly selected number) per block, at random times. Using such a task we ensured that spatial attention was drawn away from the gratings while stimulating subjects to maintain fixation, allowing us to record activity that predominantly reflected bottom-up, sensory-specific signals (Mostert et al., 2015).

Before the MEG session, the volunteers participated in a behavioral screening session that served to both train the subjects on the task as well as to assess their ability to perform it. Subjects were instructed to mentally rotate the stimulus at an angular velocity of  $30^\circ/\text{s}$  by demonstrating examples of rotations on the screen. Moreover, during this session subjects were required to press a button as soon as they achieved a vivid imagination of the grating on completion of the cued rotation. This provided a proxy of the speed at which they actually performed the rotation and was used as selection criterion for participation in the MEG session.

### Staircase procedure

The amount of jitter of the probe grating was determined online using an adaptive staircasing procedure to equalize subjective task difficulty across conditions and subjects. The starting difference was set to  $15^\circ$  and was increased by  $1^\circ$  following an incorrect response and decreased by  $0.5488$  after two consecutive correct responses. Such a procedure has a theoretical target

performance of ~80% correct (García-Pérez, 1998). Four separate staircases were used, one for each of the VWM and MR conditions.

### MEG recordings, eye-tracker recordings, and pre-processing

Neural activity was measured using a whole-head MEG system with 275 axial gradiometers (VSM/CTF Systems) situated in a magnetically shielded room. A projector outside the room projected via a mirror system onto the screen located in front of the subject. Fiducial coils positioned on the nasion and in the ears allowed for online monitoring of head position and for correction in between blocks if necessary. Both vertical and horizontal electro-oculogram (EOG) as well as electrocardiogram were obtained to aid in the recognition of artifacts. All signals were sampled at 1200 Hz and analyzed offline using the Field-Trip toolbox (Oostenveld et al., 2011). The data were notch-filtered at 50 Hz and corresponding harmonics to remove line noise, and subsequently inspected in a semi-automatic manner to identify irregular artifacts. After rejection of bad segments, independent component analysis was used to remove components that corresponded to regular artifacts such as heartbeat, blinks and eye movements (although our results suggest that the removal of eye movement-related artifacts was imperfect, see Results and Discussion). The cleaned data were baseline-corrected on the interval of -200–0 ms, relative to stimulus onset.

Gaze position and pupil dilation were continuously tracked throughout the experiment using an Eyelink 1000 (SR Researcher) eye-tracker. The eye-tracker was calibrated before each session and signals were sampled at 1200 Hz. Because we were interested in eye-movements induced by the experimental stimulation, we removed any slow drifts in the signal by baseline-correcting the signal on an interval of -200–0 ms relative to cue onset.

### Data sharing

All data, as well as analysis scripts required to obtain the presented figures, are available from the Donders Institute for Brain, Cognition, and Behavior repository at [http://hdl.handle.net/11633/di.dc-cn.DSC\\_3018016.04\\_526](http://hdl.handle.net/11633/di.dc-cn.DSC_3018016.04_526).

### Classification and decoding analyses

Originally, we first focused on the neural data. Broadly, we conducted two lines of decoding analyses. In the first, we focused only on the blocks in which participants performed the combined VWM/imagery task, using 8-fold cross-validation. We trained a three-class probabilistic classifier that returns the probability that a given trial belongs to either of the three presented grating orientations. To improve signal-to-noise ratio, yet retain the ability to draw firm conclusions regarding the timing of any decoded signal, we smoothed the data using a moving average with a window of 100 ms. The classifier was trained across the spatial dimension (i.e., using sensors as features), on trials from all conditions (i.e., all amounts and directions of rotation). This may seem counterproductive, because the mental contents diverge over the delay in-

terval and there should therefore be no systematic relationship between the MEG data and the stimulus label. However, our rationale was that regardless of condition, subjects need to first perceive, encode and maintain the presented stimulus before they can even commence the task, be it VWM or MR. Thus, we expected to be able to extract the neural pattern of the presented stimulus during at least the physical presentation and a brief moment after that. This classifier was then trained and applied across all time points, resulting in a temporal generalization matrix (King and Dehaene, 2014). It is important to note that we trained the classifiers only using the labels of the presented stimulus but sorted the data in varying ways when testing the performance. For example, by looking at an early training time point, but a late testing time point, we tested whether we could decode the orientation of the grating kept in mind near the end of the delay period, on the basis of the pattern evoked by the presented stimulus early in the trial.

In the second line of analysis, we trained a continuous orientation decoder on the functional localizer. The larger number of orientations sampled in the functional localizer allowed us to decode a continuous estimate of represented orientation, rather than a discrete one from a fixed number of classes. We applied this decoder to the VWM/imagery task and subsequently related the decoded orientation to the true presented orientation by calculating a quantity intuitively similar to a correlation coefficient (see below, Continuous orientation decoder). Here too, we extended the procedure to include all pairwise training and testing time points, resulting in temporal generalization matrices (King and Dehaene, 2014).

In the control analysis, where we tested for a systematic relationship between gaze position and VWM contents, we repeated the first line of analysis described above, but instead used the gaze position ( $x$ - and  $y$ -coordinates) as features rather than the MEG data.

### Multi-class probabilistic classifier

The three-class classifier was based on Bishop (2006, pp 196–199). Briefly, the class-conditional densities were modeled as Gaussian distributions with assumed equal covariance. By means of Bayes' theorem, and assuming a flat prior, this model was inverted to yield the posterior probabilities, given the data. Specifically, let  $\mathbf{x}$  be a column vector with length equal to the number of features [number of sensors for MEG data, two for gaze position (horizontal and vertical location)] containing the data to be classified, then the posterior probability that the data belongs to class  $k$  is given by the following equations:

$$P(\text{class} = k | \mathbf{x}) = \frac{\exp(a_k)}{\sum_j \exp(a_j)}$$

$$a_k(\mathbf{x}) = \mathbf{w}_k^T \mathbf{x} + w_{k0}$$

$$\mathbf{w}_k = \mathbf{S}^{-1} \mathbf{m}_k$$

$$w_{k0} = -\frac{1}{2} \mathbf{m}_k^T \mathbf{S}^{-1} \mathbf{m}_k$$

where  $\mathbf{m}_k$  is the mean of class  $k$  and  $\mathbf{S}$  is the common covariance, both obtained from the training set. The latter was calculated as the unweighted mean of the three covariance matrices for each individual class, and subsequently regularized using shrinkage (Blankertz et al., 2011) with a regularization parameter of 0.05 for the MEG data and 0.01 for the eye-tracker analysis.

### Continuous orientation decoder

The continuous orientation decoder was based on the forward-modeling approach as described in Brouwer and Heeger (2009, 2011) but adapted for improved performance (Kok et al., 2017). The forward model postulates that a grating with a particular orientation activates a number of hypothetical orientation channels, according to a characteristic tuning curve, that subsequently lead to the measured MEG data. We formulated a model with 24 channels spaced equally around the circle, whose tuning curves were governed by a Von Mises curve with a concentration parameter of 5. Note that all circular quantities in the analyses were multiplied by two, because the formulas we used operate on input that is periodic over a range of 360° but grating orientation only ranges from 0° to 180°. Next, we inverted the forward model to obtain an inverse model. This model reconstructs activity of the orientation channels, given some test data. In this step we departed from Brouwer and Heeger (2009, 2011)'s original formulation in two aspects. First, we estimated each channel independently from each other, allowing us to include more channels than there are stimulus classes. Second, we explicitly took into account the correlational structure of the noise, which is a prominent characteristic of MEG data, to improve decoding performance (Blankertz et al., 2011; Mostert et al., 2015). For full implementation details, see Kok et al. (2017). The decoding analysis yields a vector  $\mathbf{c}$  of length equal to number of channels (24 in our case) with the estimated channel activity in a test trial, for each pairwise training and testing time point. These channels activities were then transformed into a single orientation estimate  $\theta$  by calculating the circular mean (Berens, 2009) across all the orientations the channels are tuned for, weighted by each individual activation:

$$\theta = \arg \left[ \sum_i c_j \exp(i\mu_j) \right]$$

where the summation is over channels,  $\mu_j$  is the orientation around which the  $j$ th channel tuning curve is centered, and  $i$  is the imaginary unit. These decoded orientations can then be related to the true orientation, across trials, as follows:

$$z = \frac{1}{N} \sum_k \exp[i(\theta_k - \varphi_k)]$$

$$\rho = |z| \cos(\arg(z))$$

where  $N$  is the number of trials and  $\varphi_k$  is the true orientation on trial  $k$ . The quantity  $\rho$  is also known as the test statistic in the V test for circular uniformity, where the orientation under the alternative hypothesis is prespecified (Berens, 2009). This quantity has properties that make it intuitively similar to a correlation coefficient: it is +1 when decoded and true orientations are exactly equal, -1 when they are in perfect counterphase and 0 when there is no systematic relationship or when they are perfectly orthogonal.

### Statistical testing

All inferential statistics were performed by means of a permutation test with cluster-based multiple comparisons correction (Maris and Oostenveld, 2007). These were applied to either whole temporal generalization matrices, or horizontal cross-sections thereof (i.e., a fixed training window). These matrices/cross-sections were tested against chance-level (33%) in the classification analysis, or against zero in the continuous decoding analysis. In the first step of each permutation, clusters were defined by adjacent points that crossed a threshold of  $p < 0.05$  according to a two-tailed one-sample  $t$  test. The  $t$  values were summed within each cluster, but separately for positive and negative clusters, and the largest of these were included in the permutation distributions. A cluster in the true data were considered significant if its  $p$  value was  $< 0.05$ . For each test, 10,000 permutations were conducted.

### Spatial patterns and source analysis

To interpret the signals that the classifier and decoder pick up, we looked at the corresponding spatial patterns (Haufe et al., 2014). The spatial pattern is the signal that would be measured if the latent variable that is being decoded is varied by one unit. For both the probabilistic classifier and the continuous orientation decoder, this comes down to the difference ERF between each category and the average across all categories. This yields one spatial pattern for each class, and these were subsequently averaged across classes, as well as across time of interest, and fed into source analysis and synthetic planar gradient transformation. This transformation refers to a procedure whereby MEG data recorded with axial gradiometers is transformed as if it were measured by planar gradiometers (Bastiaansen and Knösche, 2000). The main advantage is that the spatial distribution of the resulting data is more readily interpretable.

For source analysis, we used a template anatomic scan provided by FieldTrip to create a volume conduction model based on a single shell model of the inner surface of the skull. The source model consisted of a regular grid spaced 0.5 cm apart that encompassed the entire brain. Leadfields were calculated and rank-reduced to two dimensions, to accommodate the fact that MEG is blind to tangential sources. The covariance of the data were calculated over the window of 1–8 s post-stimulus and regularized using shrinkage (Blankertz et al., 2011) with a

regularization parameter of 0.05. The leadfields and data covariance were then used to calculate linearly constrained minimum variance spatial filters (LCMV, also known as beamformers; Van Veen et al., 1997). Applying these filters to sensor-level data yields activity estimates of a two-dimensional dipole at each grid point. We further reduced these estimates to a scalar value by means of the Pythagorean theorem. This leads to a positivity bias however, that we corrected for using a permutation procedure (Manahova et al., 2017). The number of permutations was 10,000. The final result was interpolated to be projected on a cortical surface and quantifies the degree to which a particular area contributed to the performance of the classifier/decoder.

## Results

### Behavioral results

The average accuracies in the MEG session for the four conditions ranged from 68% to 72%, confirming that subjects were able to do the task, as well as that the staircase procedure was successful. The average final jitter estimate from the staircase procedure for the 0°, 60°, 120°, and 180° conditions were as follows (95%-CI in parentheses): 3.1° (0.98-5.2), 11.0° (8.87-13.0), 13.4° (11.35-15.52), and 6.5° (4.42-8.60), respectively. With the exception of the 180° condition, the rising trend in these values suggests that subjects found the task more difficult when the amount of rotation was larger. The relatively low value for the 180° condition however indicates that this condition was relatively easy. One explanation may be the fact that subjects did not require the final product of the MR to perform well on the task. It is possible that they simultaneously memorized the starting orientation. After having finished the MR, regardless of how well they were able to do so, they could simply reactivate the initial image and use that in their judgment.

### Gaze position tracks VWM contents

In the eye movement analysis, we investigated whether there is a relation between gaze position and the item held in VWM. We adopted the same analysis in our original main analysis (see Materials and Methods), but instead entered the horizontal and vertical gaze position, measured by the eye-tracker, as features in the decoding analysis. Specifically, we constructed a three-class probabilistic classifier that yields the posterior probabilities that any given data belong to either of three presented orientations. That is, the classifier was trained according to the labels of the presented stimulus. The classifier was trained on trials from all conditions (i.e., all amounts and directions of rotation) pooled together to obtain maximum sensitivity (for rationale, see Materials and Methods). To verify whether we could decode stimulus identity from the gaze position, we first applied the classifier to the same (pooled) data using cross-validation. We found above-chance decoding in a time period of ~0.5–3.5 s post-stimulus that was marginally significant (Fig. 2, Extended Data Fig. 2-1). This indicates that subjects moved their eyes in a way consistent with the present stimulus, and kept it there for ~2–3 s.

Then, when looking at the decoded signal within the VWM condition only, we found a sustained pattern (Fig. 2A), although again only marginally significant. This suggests that on perceiving and encoding the stimulus, subjects move their eyes in a way systematically related to the identity of the stimulus and keep that gaze position stable throughout the entire delay period.

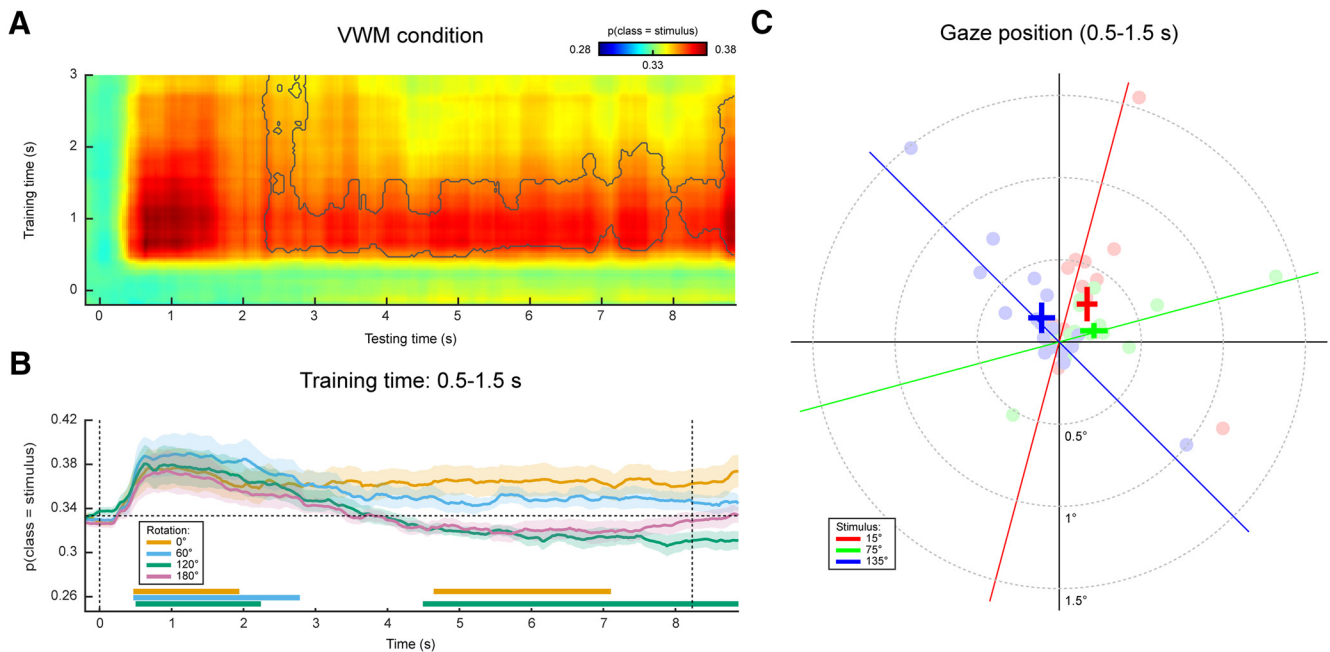
Contrary to previously used paradigms, where two stimuli were displayed at the beginning of a trial and a retro-cue signaled the item that was to be remembered (Harrison and Tong, 2009; Albers et al., 2013; Christophel et al., 2015, 2017), in the present experiment we only showed one stimulus. It is therefore possible that the sustained decoding performance does not necessarily reflect VWM contents, but simply that the subjects moved their gaze according to the presented stimulus rather than to their mental contents. However, if this were true, then we should find a similar effect in the three MR conditions. If, on the other hand, the classifier picked up the item kept in mind, then the probability that a trial is assigned to the same class as the presented stimulus should drop over time, as the subject rotates the mentally imagined grating away from the starting orientation. Our results were consistent with the latter scenario (Fig. 2B). Whereas the probability that the data belong to the same class as the presented stimulus stays steadily above chance in the VWM condition, it drops to lower levels in the three MR conditions. Moreover, we found evidence that the gaze moves toward a position consistent with the orientation of the presented grating plus or minus 60° (depending on the cued direction of rotation) in the MR conditions, but not any further (Fig. 3A,B).

Figure 2C displays the grand average, as well as individual average gaze positions during 0.5–1.5 s after stimulus onset, separately for each of the three stimulus conditions, collapsed across VWM and MR conditions. Although there is large variability among subjects in the magnitude of the eye movements, a general trend can be discerned where subjects position their gaze along the orientation axis of the grating. The mean disparity in visual angle with respect to pre-trial fixation was only 0.23°, which is in the same order of magnitude as reported previously (Foster et al., 2016), although for some subjects, it was larger, up to 1.5°.

In short, there was a systematic relationship between gaze position and stimulus orientation, after which the gaze position tracked the orientation kept in mind during the delay period, but only for a maximum of approximately  $\pm 60^\circ$  relative to starting orientation. These findings raise the concern that any potential decoding of VWM items from MEG signals, as was the aim of our original analysis, could be the result of stimulus-related eye confounds (for possible underlying mechanisms, see Discussion).

### Sustained decoding of VWM items from MEG signals

The original aim of this study was to assess the representational contents of the neural signals while the subjects were engaged in VWM/imagery. We present these results here, to demonstrate how they could easily have been mistaken for genuine results, had we been oblivious



**Figure 2.** Gaze position classification results, cross-validation within VWM/imagery task. **A**, Temporal generalization matrix of classification performance in the VWM condition. The color scale denotes the average posterior probability that the data belong to the same class as the presented stimulus. The gray outline demarcates a near-significant cluster ( $p = 0.069$ ). Note that this matrix is asymmetric because only the VWM condition is shown, while the classifier was trained on the data from all VWM/MR conditions. For this reason, the data after  $\sim 3$  s are not expected to contain systematic patterns and therefore the training time axis has been truncated (Extended Data Fig. 2-1). **B**, Classification performance averaged over the training time window of 0.5–1.5 s, separately for the VWM and the three MR conditions. Note that the  $0^\circ$  condition corresponds directly to the matrix in **A**. The two vertical dashed lines indicate stimulus and probe onset. Shaded areas indicate the SEM. Significant clusters are indicated by the horizontal bars in the lower part of the figure. **C**, Average gaze position during 0.5–1.5 s after stimulus onset, separately per stimulus orientation. Each translucent dot corresponds to an individual subject. The crosses are the grand averages, where the vertical and horizontal arms denote the SEM. The three colored lines depict the orientation of the three stimuli.

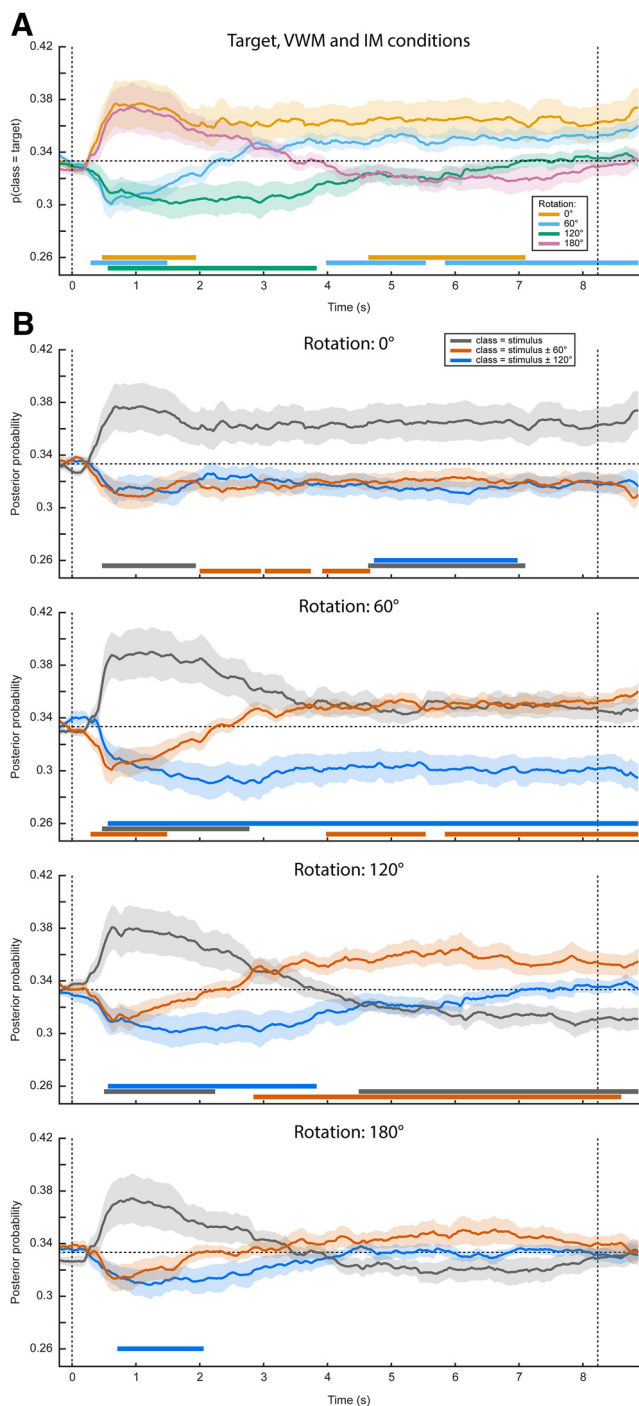
to the systematic eye movements. We constructed a three-class probabilistic classifier in which the MEG sensors were entered as features. As before, the classifier was trained on trials from all conditions (i.e., all amounts and directions of rotation) pooled together for maximum sensitivity (for rationale, see Materials and Methods). To test whether we could decode stimulus identity from the MEG signal, we applied the classifier to the same (pooled) data using cross-validation and found successful decoding during a period of up to  $\sim 2.5$  s after stimulus onset (Fig. 4, Extended Data Fig. 4-1). The stimulus itself was presented for only 250 ms. Therefore, the later part of this period could have been interpreted as an endogenous representation, for instance stemming from active mental instantiation by the subject, although in reality it is more likely to be the result of eye movements.

Next, we looked at the decoding performance in the VWM condition alone, using the classifier trained on all conditions as described above. We found that the identity of the memory item could be decoded during the entire delay interval, using classifiers obtained from a training window of  $\sim 0.5$ –1.5 s (Fig. 4A). The performance stayed above chance-level (33.3%) at a stable level of  $\sim 37\%$  throughout the entire interval (Fig. 4B).

Again, we found this sustained pattern to be specific to the VWM condition, because the probability that the data belong to the same class as the presented stimulus drops

over time in the three MR conditions (Fig. 4B). As explained in the previous paragraph, this indicates that the sustained above-chance classification in the VWM condition cannot be explained as a long-lasting stimulus-driven effect (e.g., stimulus aftereffect), but must also reflect the memorized item to at least some degree. In the  $180^\circ$  condition, the posterior probability that the data belong to the same classes as the presented stimulus later re-emerges as a rising, although nonsignificant trend. This can be explained by the fact that the final orientation that the subjects should have in mind in the  $180^\circ$  condition is identical to the orientation of the presented stimulus at the start of a trial.

To facilitate interpretation of these results, we inspected the classifier's corresponding sensor topography (Fig. 4C) and source localization (Fig. 4D), averaged over the training time period of 0.5–1.5 s. These indicate that both occipital (Harrison and Tong, 2009; Albers et al., 2013) and prefrontal sources (Sreenivasan et al., 2014; Spaak et al., 2017) contributed to the classifier's performance. Indeed, the prefrontal sources could in reality point to ocular sources. Moreover, although the contribution from occipital regions may seem to provide evidence that the decoder genuinely picks up visual representations, these sources could in fact also be driven by the eye movements (see Discussion).



**Figure 3.** Complete gaze position classification results. Analogously to [Figure 2](#), the classifier was trained on the time window of 0.5–1.5 s after stimulus onset, and according to the labels of the presented stimulus. **A**, Average posterior probability that the data belong to the class of the target orientation, that is, the orientation that the subjects were supposed to have in mind at the end of the delay period. For both the 0° and the 180° conditions, the target orientation was the same as the presented stimulus. For the 60° and 120° conditions, however, the target and presented stimulus were different, hence the below-chance probabilities at the beginning of the delay period. **B**, The average posterior probabilities that the data belong to either of three classes: the same orientation as the presented stimulus, the

### Figure 3. continued

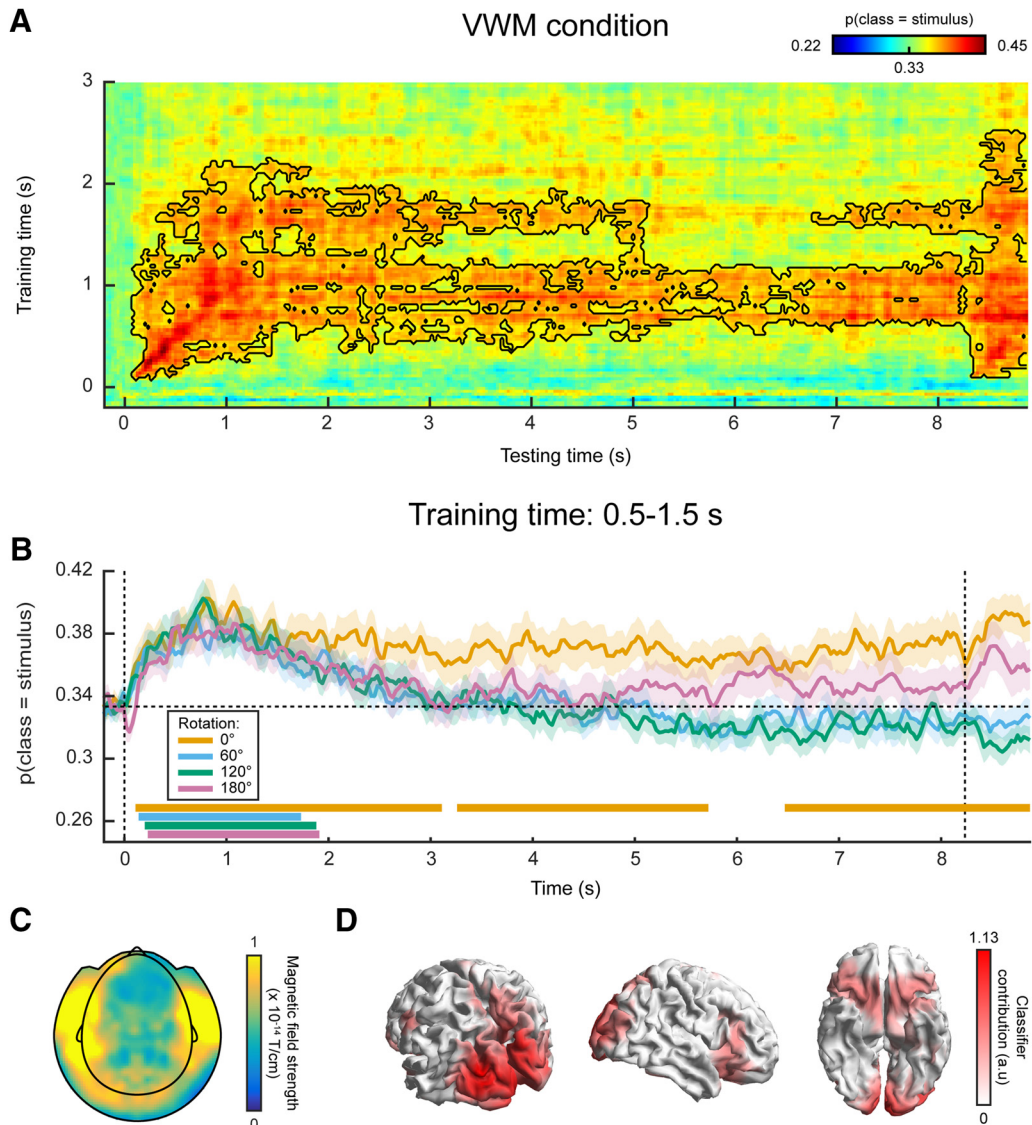
orientation of the presented stimulus  $\pm 60^\circ$  or the orientation of the presented stimulus  $\pm 120^\circ$ , plotted separately for the four VWM/MR conditions. The plus/minus sign is due to the MR being performed either clockwise or counterclockwise. This figure gives insight into whether the feature patterns corresponding to any intermediate orientations become active during MR, which is particularly relevant for the 120° and 180° conditions. For instance, in the 180° counterclockwise condition, the subject would start with a mental image with the same orientation as the presented stimulus, then pass through, respectively,  $-60^\circ$  and  $-120^\circ$ , ultimately to reach the target of  $-180^\circ$  (i.e.,  $0^\circ$ ). If the gaze positions corresponding to all these orientations become active in sequence, one would first expect a peak in posterior probability that the data belong to the same class as the stimulus (gray line, bottom figure), then a peak in the probability of belonging to the presented stimulus  $-60^\circ$  (orange line), then for  $-120^\circ$  (blue line), and finally again for  $0^\circ$  (gray line). It is important to realize that below-chance probabilities in these analyses are meaningful. For instance, consider **A**. Here, the probability that the data belong to the same class as the target is plotted. Hence, in the 60° and 120° conditions, the classifier correctly identifies that the data do not belong to the same class as the target in the beginning of the interval, because the starting orientation was different. As another example, consider the red line in the second panel in **B**. This line plots the probability that the data belong to the starting orientation  $\pm 60^\circ$ . On presentation of the starting orientation, the classifier therefore yields a significant below-chance probability. However, as the subject performs the  $\pm 60^\circ$  rotation over the course of the trial, the classifier increasingly picks up this rotated image, hence giving above-chance probabilities. Note that **A**, **B**, as well as [Figure 2](#), all depict the same data but visualized in different manners. Shaded areas denote the SEM and significant clusters are depicted by the thick horizontal lines at the bottom of the panels.

Finally, we investigated whether we could decode the intermediate (for the 120° and 180° rotations) and the final orientations in the MR conditions. We found some indication that the final orientation, but not the intermediate ones ([Fig. 5B](#)), indeed emerges halfway through the delay period, but this effect was not statistically significant ([Fig. 5A](#)).

### Decoding visual representations from sensory areas

In a third analysis, we trained a continuous orientation decoder (see Materials and Methods) on the functional localizer data ([Fig. 1B,C](#)) and applied these decoders to the data from the VWM/imagery task ([King and Dehaene, 2014](#)). The main advantage of this method is that it ensures that the decoder is primarily sensitive to sensory signals, and not to higher-level top-down processes involved in mental manipulation of an image. It thus allows us to track sensory-specific activation throughout the delay period ([Mostert et al., 2015](#)). Cross-validation within the functional localizer confirmed that we were indeed able to reliably decode orientation-specific information from activity evoked by passively perceived gratings ([Fig. 6](#), Extended Data [Fig. 6-1](#)). Moreover, we were not able to decode grating orientation on the basis of gaze position, verifying that the data from the functional localizer were not contaminated by stimulus-specific eye movements (Extended Data [Fig. 6-2](#)).





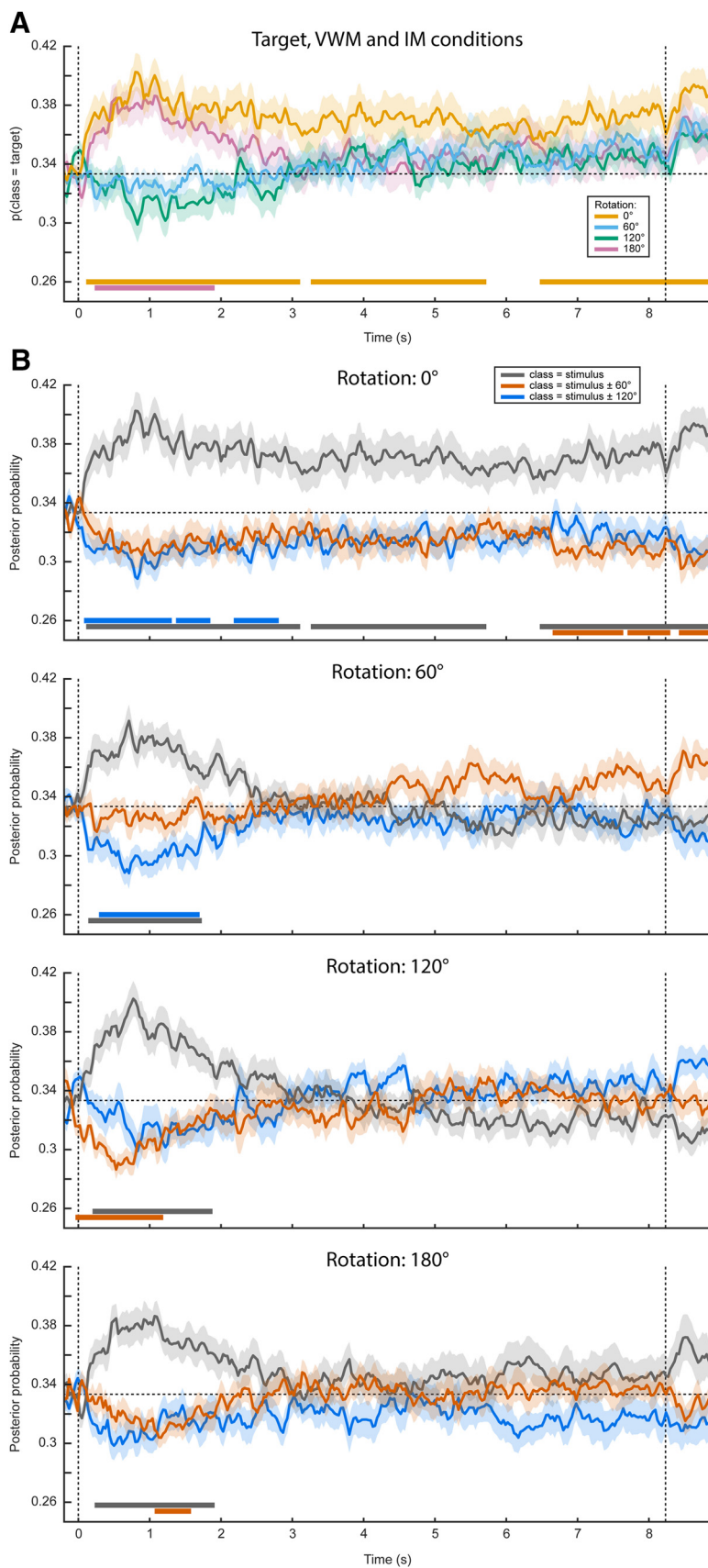
**Figure 4.** MEG classification results. **A**, Same as in Figure 2A, except the classification was performed on MEG data, rather than on gaze position. The black outline demarcates a significant cluster ( $p = 0.006$ ). **B**, Same as in Figure 2B, except the analysis was performed on MEG data. Synthetic planar gradiometer topography (**C**) and source topography (**D**) of areas that contributed to the classifier. See also Extended Data Figure 4-1.

In the VWM condition, the decoders trained on the functional localizer data could reliably decode the orientation of the presented stimulus (Fig. 6A). Moreover, for a training time of ~90–120 ms, we could decode the stimulus for a prolonged time, lasting over 1 s after stimulus onset. Interestingly, this training time point coincides with the time at which peak performance is obtained within the localizer itself (Extended Data Fig. 6-1). Comparing the decoding trace within this training time window with the three MR conditions, it can be seen that grating orientation can be decoded in all four conditions for a sustained period of ~500 ms (Fig. 6B). This is in sharp contrast to the extended decoding of the presented stimulus throughout the entire delay period, found within the VWM/imagery task using cross-validation (Fig. 4).

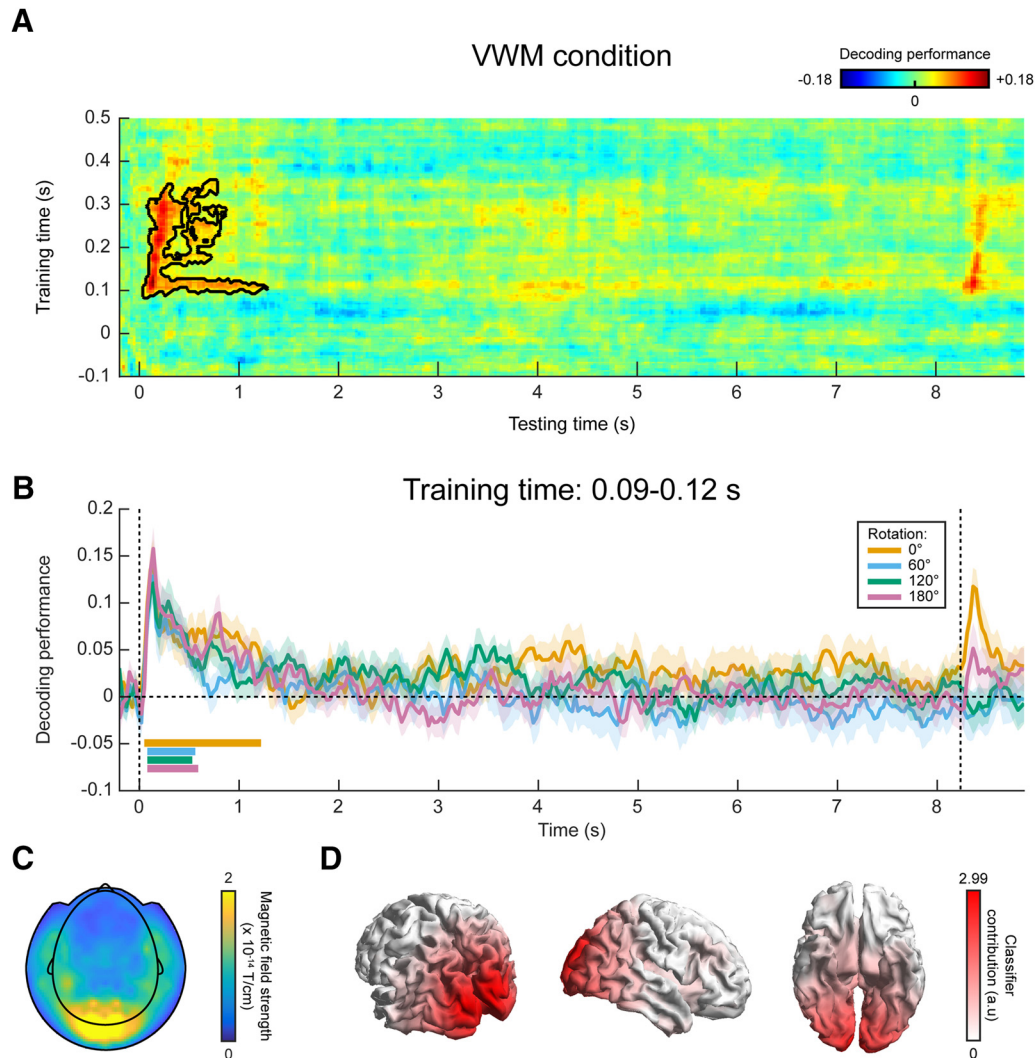
We inspected the spatial pattern (Fig. 6C) and corresponding source topography (Fig. 6D) for the decoders,

averaged across training time 90–120 ms. These highlight primarily occipital regions as contributing to the decoder’s performance, consistent with our premise that the functional localizer primarily induced bottom-up sensory signals, especially during this early time interval (Mostert et al., 2015).

In summary, our findings suggest that the stable, persistent representation found in our within-task MEG decoding result may well be attributed to stimulus-specific eye movements, although the magnitude of the eye movements were only small. In contrast, no clear evidence was found for such a long-lasting representation when training the decoder on the functional localizer. Given that the localizer was not contaminated by stimulus-specific eye movements, these results thus provide a more reliable picture of the sensory representations during the delay interval.



**Figure 5.** Complete MEG classification results, visualized in a variety of ways. This figure is analogous to Figure 3, except the classifier is trained and tested on MEG data rather than on gaze position.



**Figure 6.** MEG decoding results, generalized from localizer to VWM/imagery task. **A**, Temporal generalization matrix of orientation decoding performance, for which the decoder was trained on all time points in the functional localizer and tested across all time points in the VWM/imagery task. The color scale reflects the correspondence between true and decoded orientation. The black outline shows a significant cluster ( $p = 0.04$ ). Note that the x- and y-axis in the figure are differently scaled for optimal visualization. **B**, Decoding performance over time in the VWM/imagery task, averaged over decoders trained in the window of 0.09–0.12 s in the localizer, separately for the VWM and three MR conditions. Shaded areas denote the SEM, and significant clusters are indicated by the horizontal bars. Synthetic planar gradiometer topography (**C**) and source topography (**D**) of areas that contribute to the decoder. See also Extended Data Figures 6-1, 6-2.

## Discussion

Neural decoding is a powerful and promising technique for neuroimaging studies (Haxby et al., 2014; Grootswagers et al., 2017) that has led to substantial advancement in the field of VWM over recent years (Harrison and Tong, 2009; Serences et al., 2009; Albers et al., 2013; Wolff et al., 2015, 2017). This study also employed neural decoding techniques, with the original aim to investigate the temporal dynamics of sensory representations during VWM. However, we found that our data were contaminated by small but systematic eye movements, whereby gaze position was related to the stimulus held in mind. This jeopardized our ability to interpret the neural decoding results, results that otherwise would have seemed

sensible, because very similar results could be obtained by considering gaze position only.

There are at least three possible mechanisms via which stimulus-specific eye movements may confound our results. First, eye movements are known to cause stereotypical artifacts in MEG recordings. Due to the positively charged cornea and negatively charged retina, the eyeball acts as an electromagnetic dipole whose rotation is picked up by the MEG sensors. The spatial pattern that the dipole evokes is directly related to its angle, or in other words, to the position of the subject's gaze (Plöchl et al., 2012). Thus, if the subject moves his/her eyes in response to the grating in a manner related to the orientation of that grating, then this will induce a specific pattern in the MEG signals, which in turn is directly related to the grating

orientation. A decoding analysis applied to these signals is then likely to pick up the patterns evoked by the eyeball dipoles, confounding potential orientation-related information stemming from genuine neural sources. In fact, our source analysis hints at this scenario (Fig. 4D), as the contributions from presumed prefrontal sources closely resemble an ocular source.

Second, if the eyes move, then the projection falling on the retina will also change, even when external visual stimulation remains identical. Thus, if gaze position is systematically modulated by the image that is perceived or kept in mind, then so is the visual information transmitted to the visual cortex. For example, if a vertical grating is presented and kept in VWM, then the subject may subtly move her or his gaze upward. Correspondingly, the fixation dot is now slightly below fixation, thus leading to visual cortex activity that is directly related to the retinotopic position of the fixation dot. Our decoding analysis may thus actually decode the position of the fixation dot, rather than grating orientation, potentially leading to an incorrect conclusion. Source analysis would in this scenario also point to occipital sources, similarly to what we found (Fig. 4D). Note that this mechanism is not specifically dependent on the presence of a fixation dot. A systematic difference in eye position will also lead to changes in the retinotopic position of, for instance, the presentation display or the optically visible part of the MEG helmet.

Third, if gaze position covaries with the mental image, then decoding of the mental image will also reveal areas that encode eye gaze position, such as oculomotor regions in parietal and prefrontal cortex.

Our findings raise the question of why there were task-induced eye movements that were directly related to the grating kept in VWM. In fact, there is a considerable mass of literature that describes the role of eye movements in mental imagery. It has been found that subjects tend to make similar eye movements during imagery as during perception of the same stimulus (Brandt and Stark, 1997; Laeng and Teodorescu, 2002; Laeng et al., 2014). Already proposed by Donald Hebb (Hebb, 1968), it is now thought that eye movements serve to guide the mental reconstruction of an imagined stimulus, possibly by dwelling on salient parts of the image (Spivey and Geng, 2001; Laeng et al., 2014). Moreover, the specificity of the eye movements is also related to neural reactivation (Bone et al., 2017) and recall accuracy (Laeng and Teodorescu, 2002; Laeng et al., 2014; Bone et al., 2017). Our findings are in accordance with these studies. Subjects' gaze was positioned along the orientation axis of the grating - that is, the visual location within the stimulus that provided the highest information regarding its orientation and is thus exactly what one would expect given that the task was to make a fine-grained orientation comparison with a probe grating. Importantly, however, subjects were explicitly instructed to maintain fixation throughout the entire trial. We nevertheless observed that not all subjects adhered to this requirement, albeit involuntarily.

Despite these problems associated with the systematic eye movements in our experiment, it is still possible that

our decoding results do in reality stem from genuine orientation information encoded in true neural sources. In fact, we used independent component analysis in our pre-processing pipeline to (presumably) remove eye-movement artifacts. However, it would be very difficult, if not impossible, to convincingly establish that no artifacts remain and, considering the similarities between the decoding results from the MEG data (Fig. 4A,B) and the gaze position (Fig. 2A,B), we feel any attempts at this would be unwarranted.

Given the potential pervasiveness of systematic eye movements in VWM/imagery tasks, and the demonstrated susceptibility of our analysis methods to these confounds, one wonders whether other studies may have been similarly affected. Clearly, the first mechanism described above involving the eyeball dipole would only affect electrophysiological measurements like electroencephalography and MEG, and has indeed been a concern in practice (Foster et al., 2016). The second mechanism however, whereby stimulus identity is confounded with the retinal position of visual input, would also affect other neuroimaging techniques such as fMRI. This confound could be particularly difficult to recognize, because it would also affect activity in visual areas. Moreover, because eye movements during imagery have been found to be positively related to performance (Laeng and Teodorescu, 2002), this could potentially explain correlations between VWM decoding and behavioral performance. The third mechanism, whereby one directly decodes gaze position from motor areas, could be a problem especially for fMRI which, thanks to its high spatial resolution, might be well able to decode such subtle neural signals. This concern may be especially relevant for studies that investigate the role of areas involved in eye movements or planning thereof, such as frontal eye fields or superior precentral sulcus, in the maintenance of working memory items (Jerde et al., 2012; Ester et al., 2015; Christophel et al., 2017).

This leaves the question of how to deal with eye movements in VWM/imagery tasks. Naturally, it is important to record eye movements during the experiment, for instance using an eye-tracker or EOG. One can then test for any systematic relationship and, if found, investigate whether it could confound the main results. In our case, for example, decoding of gaze position leads to strikingly similar results as those obtained from the MEG data. Foster et al. (2016) on the other hand found that decoding performance of working memory items decreased throughout the trial, whereas the deviation in gaze position increased, suggesting that eye confounds cannot explain the main findings. Another approach might be to design the experimental task in such a way that eye movements are less likely. For example, by presenting gratings laterally (Pratte and Tong, 2014; Ester et al., 2015; Wolff et al., 2017), and assuming that VWM items are stored in a retinotopically specific manner (Pratte and Tong, 2014), the involuntary tendency to move one's eyes subtly along the remembered grating's orientation axis may become less strong, because those gratings are located distantly from the gaze's initial location (i.e., cen-

tral fixation). Finally, a powerful approach could be to adopt a separate functional localizer, which allows specific decoding of functionally defined representations such as bottom-up, sensory-specific signals (Harrison and Tong, 2009; Serences et al., 2009; Albers et al., 2013; Mostert et al., 2015). If the localizer is well designed and not systematically contaminated by eye movements, then eye movements in the main task cannot have a systematic effect on the decoded signal, thus effectively filtering them out.

We designed a localizer that is specifically sensitive to the neural representations encoded in bottom-up signals evoked by passively perceived gratings. This allowed us to address the question of whether the imagined stimulus was encoded with a similar neural code as the perceived gratings (Harrison and Tong, 2009; Albers et al., 2013). Using this localizer, we indeed obtained MEG decoding results that were very dissimilar from those obtained using cross-validation within the combined VWM/imagery task. We no longer found persistent activation of an orientation-specific representation throughout the entire delay period. Nevertheless, the sensory pattern did remain above baseline for a period of ~1 s, which is relatively long considering that the stimulus was presented for only 250 ms. One explanation is that the stimulus was relevant for the task. Previous work has shown that task relevance may keep the sensory representation online for a prolonged period even after the stimulus is no longer on the screen (Mostert et al., 2015).

It should be pointed out that using a functional localizer also has its intrinsic limitations. The most important being that, while such an approach is primarily sensitive to a functionally defined signal, it may at the same time be blind to other relevant signals that were not a priori included in the functional definition. The VWM literature itself provides an instructive example: while the functional localizer approach has clearly demonstrated sensory representations of the memorandum in associated sensory cortex (Harrison and Tong, 2009; Albers et al., 2013), it would have missed relevant encoding in other regions in the brain such as parietal and prefrontal cortex (Christophel et al., 2015, 2017). Furthermore, the fact that the decoders were trained on a functionally defined signal does not mean that they are necessarily insensitive to other signals, such as eye movement-related signals. However, it is important to realize that the exact effect of these other signals on the decoder's output is not explicitly defined. These potential effects would therefore be idiosyncratic to an individual's data and are expected to cancel out at the group level.

In summary, we demonstrate a case where decoding analyses in a VWM/imagery task are heavily confounded by systematic eye movements. Given the high potential benefit of decoding analyses and its widespread use in the study of working memory and mental imagery, we argue that this problem may be more pervasive than is commonly appreciated. Future studies could target this question specifically and investigate how strong the confounds are exactly. One approach could be to systematically vary salient input and assess how this impacts

decoding performance (cf. the second mechanism described above). Furthermore, it is important to realize that this does not necessarily invalidate all previous studies. While some previous results may have been afflicted, our current understanding of the neural underpinnings of VWM is still firmly grounded in converging evidence from a wide variety of techniques, paradigms and modalities. Nevertheless, we conclude that eye movement confounds should be taken seriously in both the design as well as the analysis phase of future studies.

## References

- Albers AM, Kok P, Toni I, Dijkerman HC, de Lange FP (2013) Shared representations for working memory and mental imagery in early visual cortex. *Curr Biol* 23:1427–1431. [CrossRef Medline](#)
- Bastiaansen MCM, Knösche TR (2000) Tangential derivative mapping of axial MEG applied to event-related desynchronization research. *Clin Neurophysiol* 111:1300–1305. [CrossRef](#)
- Berens P (2009) CircStat: a MATLAB toolbox for circular statistics. *J Stat Softw* 31. Available at <https://www.jstatsoft.org/v031/i10>. [CrossRef](#)
- Bishop CM (2006) *Pattern recognition and machine learning*. New York: Springer.
- Blankertz B, Lemm S, Treder M, Haufe S, Müller K-R (2011) Single-trial analysis and classification of ERP components—a tutorial. *Neuroimage* 56:814–825. [CrossRef Medline](#)
- Bone MB, St-Laurent M, Dang C, McQuiggan DA, Ryan JD, Buchsbaum BR (2017) Eye-movement reinstatement and neural reactivation during mental imagery. *bioRxiv* 107953.
- Brandt SA, Stark LW (1997) Spontaneous eye movements during visual imagery reflect the content of the visual scene. *J Cogn Neurosci* 9:27–38. [CrossRef Medline](#)
- Brouwer GJ, Heeger DJ (2009) Decoding and reconstructing color from responses in human visual cortex. *J Neurosci* 29:13992–14003. [CrossRef Medline](#)
- Brouwer GJ, Heeger DJ (2011) Cross-orientation suppression in human visual cortex. *J Neurophysiol* 106:2108–2119. [CrossRef Medline](#)
- Christophel TB, Cichy RM, Hebart MN, Haynes JD (2015) Parietal and early visual cortices encode working memory content across mental transformations. *Neuroimage* 106:198–206. [CrossRef Medline](#)
- Christophel TB, Allefeld C, Endisch C, Haynes JD (2017) View-independent working memory representations of artificial shapes in prefrontal and posterior regions of the human brain. *Cereb Cortex* 28:2146–2161.
- Ester EF, Sprague TC, Serences JT (2015) Parietal and frontal cortex encode stimulus-specific mnemonic representations during visual working memory. *Neuron* 87:893–905. [CrossRef Medline](#)
- Foster JJ, Sutterer DW, Serences JT, Vogel EK, Awh E (2016) The topography of alpha-band activity tracks the content of spatial working memory. *J Neurophysiol* 115:168–177. [CrossRef Medline](#)
- García-Pérez MA (1998) Forced-choice staircases with fixed step sizes: asymptotic and small-sample properties. *Vision Res* 38:1861–1881. [CrossRef](#)
- Gayet S, Guggenmos M, Christophel TB, Haynes J-D, Paffen CLE, van der Stigchel S, Sterzer P (2017) Visual working memory enhances the neural response to matching visual input. *J Neurosci* 37:6638–6647. [CrossRef](#)
- Grootswagers T, Wardle SG, Carlson TA (2017) Decoding dynamic brain patterns from evoked responses: a tutorial on multivariate pattern analysis applied to time series neuroimaging data. *J Cogn Neurosci* 29:677–697. [CrossRef](#)
- Harrison SA, Tong F (2009) Decoding reveals the contents of visual working memory in early visual areas. *Nature* 458:632–635. [Cross-Ref Medline](#)
- Haufe S, Meinecke F, Görgen K, Dähne S, Haynes J-D, Blankertz B, Bießmann F (2014) On the interpretation of weight vectors of linear

- models in multivariate neuroimaging. *Neuroimage* 87:96–110. [CrossRef Medline](#)
- Haxby JV, Connolly AC, Guntupalli JS (2014) Decoding neural representational spaces using multivariate pattern analysis. *Annu Rev Neurosci* 37:435–456. [CrossRef Medline](#)
- Hebb DO (1968) Concerning imagery. In: *Images, perception, and knowledge*, pp 139–153. Dordrecht: Springer.
- Jerde TA, Merriam EP, Riggall AC, Hedges JH, Curtis CE (2012) Prioritized maps of space in human frontoparietal cortex. *J Neurosci* 32:17382–17390. [CrossRef Medline](#)
- King JR, Dehaene S (2014) Characterizing the dynamics of mental representations: the temporal generalization method. *Trends Cogn Sci* 18:203–210. [CrossRef Medline](#)
- King JR, Pescetelli N, Dehaene S (2016) Brain mechanisms underlying the brief maintenance of seen and unseen sensory information. *Neuron* 92:1122–1134. [CrossRef Medline](#)
- Kleiner M, Brainard D, Pelli D, Ingling A, Murray R, Broussard C (2007) What's new in Psychtoolbox-3. *Perception* 36:1.
- Kok P, Mostert P, de Lange FP (2017) Prior expectations induce prestimulus sensory templates. *Proc Natl Acad Sci USA* 114:10473–10478.
- Laeng B, Teodorescu DS (2002) Eye scanpaths during visual imagery reenact those of perception of the same visual scene. *Cogn Sci* 26:207–231. [CrossRef](#)
- Laeng B, Bloem IM, D'Ascenzo S, Tommasi L (2014) Scrutinizing visual images: the role of gaze in mental imagery and memory. *Cognition* 131:263–283. [CrossRef Medline](#)
- Manahova ME, Mostert P, Kok P, Schoffelen J-M, de Lange FP (2017) Stimulus familiarity and expectation jointly modulate neural activity in the visual ventral stream. *bioRxiv* 192518.
- Maris E, Oostenveld R (2007) Nonparametric statistical testing of EEG- and MEG-data. *J Neurosci Methods* 164:177–190. [CrossRef Medline](#)
- Mostert P, Kok P, de Lange FP (2015) Dissociating sensory from decision processes in human perceptual decision making. *Sci Rep* 5:18253. [CrossRef Medline](#)
- 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* 2011:156869. [CrossRef](#)
- Plöchl M, Ossandón JP, König P (2012) Combining EEG and eye tracking: identification, characterization, and correction of eye movement artifacts in electroencephalographic data. *Front Hum Neurosci* 6:278. [CrossRef](#)
- Pratte MS, Tong F (2014) Spatial specificity of working memory representations in the early visual cortex. *J Vis* 14:22. [CrossRef Medline](#)
- Rademaker RL, Serences JT (2017) Pinging the brain to reveal hidden memories. *Nat Neurosci* 20:767–769. [CrossRef Medline](#)
- Rose NS, LaRocque JJ, Riggall AC, Gosseries O, Starrett MJ, Meyerling EE, Postle BR (2016) Reactivation of latent working memories with transcranial magnetic stimulation. *Science* 354:1136–1139. [CrossRef Medline](#)
- Serences JT, Ester EF, Vogel EK, Awh E (2009) Stimulus-specific delay activity in human primary visual cortex. *Psychol Sci* 20:207–214. [CrossRef Medline](#)
- Spaak E, Watanabe K, Funahashi S, Stokes MG (2017) Stable and dynamic coding for working memory in primate prefrontal cortex. *J Neurosci* 37:6503–6516. [CrossRef Medline](#)
- Spivey MJ, Geng JJ (2001) Oculomotor mechanisms activated by imagery and memory: eye movements to absent objects. *Psychol Res* 65:235–241. [Medline](#)
- Sreenivasan KK, Curtis CE, D'Esposito M (2014) Revisiting the role of persistent neural activity during working memory. *Trends Cogn Sci* 18:82–89. [CrossRef](#)
- Stokes MG (2015) 'Activity-silent' working memory in prefrontal cortex: a dynamic coding framework. *Trends Cogn Sci* 19:394–405. [CrossRef Medline](#)
- Stokes MG, Kusunoki M, Sigala N, Nili H, Gaffan D, Duncan J (2013) Dynamic coding for cognitive control in prefrontal cortex. *Neuron* 78:364–375. [CrossRef Medline](#)
- Van Veen BD, van Drongelen W, Yuchtman M, Suzuki A (1997) Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. *IEEE Trans Biomed Eng* 44:867–880. [CrossRef Medline](#)
- Wolff MJ, Ding J, Myers NE, Stokes MG (2015) Revealing hidden states in visual working memory using electroencephalography. *Front Syst Neurosci* 9:123.
- Wolff MJ, Jochim J, Akyürek EG, Stokes MG (2017) Dynamic hidden states underlying working-memory-guided behavior. *Nat Neurosci* 20:864–871.