Occipital, parietal, and frontal cortices selectively maintain task-relevant features of multi-feature objects in visual working memory

Qing Yu, Won Mok Shim

Department of Psychological and Brain Sciences, Dartmouth College, Hanover, NH 03755, USA
Department of Biomedical Engineering, Sungkyunkwan University (SKKU), Suwon 16419, Republic of Korea
Center for Neuroscience Imaging Research, Institute for Basic Science (IBS), Suwon 16419, Republic of Korea

A B S T R A C T

Previous studies have shown that information held in visual working memory is represented in the occipital, parietal, and frontal cortices. However, less is known about whether the mnemonic information of multi-feature objects is modulated by task demand in the parietal and frontal regions. To address this question, we asked participants to remember either color or orientation of one of the two colored gratings for a delay. Using fMRI and an inverted encoding model, we reconstructed population-level, feature-selective responses in the occipital, parietal and frontal cortices during memory maintenance. We found that not only orientation but also color information can be maintained in higher-order parietal and frontal cortices as well as the early visual cortex when it was cued to be remembered. Conversely, neither the task-irrelevant feature of the cued object, nor any feature of the uncued object was maintained in the occipital, parietal, or frontal cortices. These results suggest a highly selective mechanism of visual working memory that maintains task-relevant features only.

Introduction

Visual working memory (VWM) refers to the ability to maintain and manipulate visual information in mind for a brief period of time when it is no longer in view (Curtis and D’Esposito, 2003; Khan and Muly, 2011; Luck and Vogel, 2013). As humans have highly limited cognitive resources, VWM performance is largely influenced by the ability to selectively attend to, memorize, and manipulate information according to task demand. The parietal and frontal cortices have long been considered to play a role in these processes, such as selecting task-relevant information, and ignoring or even suppressing task-irrelevant information, in order to prioritize the current task goals and optimize the deployment of limited cognitive resources (D’Esposito et al., 2000; Miller and Cohen, 2001).

In the neurophysiology literature on non-human primates, there has been ample evidence showing that parietal and frontal neurons can exhibit feature selectivity when the feature is task-relevant (Bichot et al., 1996; Buschman et al., 2012; Toth and Assad, 2002). More recent studies have also shown that such feature-selective tuning can be fine-scaled and modulated by task demand (Ibos and Freedman, 2014). In addition, human neuroimaging work using multi-voxel pattern analysis (MVPA) has shown that the parietal (Albers et al., 2013; Bettencourt and Xu, 2016; Christophel et al., 2012) and frontal (Albers et al., 2013; Lee et al., 2013) cortices, as well as the early visual cortex (Harrison and Tong, 2009; Serences et al., 2009), can represent the feature-specific information held in VWM. Furthermore, recent work using an encoding technique showed that feature-specific representations can be reconstructed at the population-level from activities in some of the parietal and frontal regions during VWM maintenance (Ester et al., 2015). This work has provided a novel tool of examining feature-selective responses during VWM maintenance in higher-level cortical areas.

However, it remains to be examined if these feature-specific representations can reflect the selective maintenance of task-relevant information, especially when multiple features are presented simultaneously. Although it has been previously shown that manipulation of attention to visual features can modulate feature-selective signals in the early visual cortex (Jhee et al., 2011) as well as in the parietal and frontal cortices (Ester et al., 2016; Liu et al., 2011) during visual attention tasks, few studies have examined how the feature-selective mnemonic representations are modulated by task demand in the parietal and frontal cortices (for early visual cortex see Serences et al., 2009).

In the current study, we aimed to systematically investigate the
mnemonic representations of both task-relevant and task-irrelevant information in the occipital, parietal, and frontal cortices when multi-feature objects were in view. We measured participants’ BOLD activity using fMRI while they performed a delay-estimation task on one of the visual features (color or orientation) of multi-feature objects. We examined mnemonic representations of color in particular because unlike orientation, color does not contain any spatial component; therefore, the memory-related signals in the parietal and frontal cortices are less influenced by spatial attention or motor-related signals which have also been implicated in frontoparietal regions (Kastner et al., 2007; Schluppeck et al., 2005; Silver et al., 2005). Using an inverted encoding model (Brouwer and Heeger, 2009, 2011; Ester et al., 2007; Schluppeck et al., 2005; Silver et al., 2005). Using an inverted encoding model (Brouwer and Heeger, 2009, 2011; Ester et al., 2007; Schluppeck et al., 2005; Silver et al., 2005), we reconstructed population-level feature representations during the delay phase. Our results revealed that VWM representations in the occipital, parietal, and frontal cortices are strongly modulated by task, and that only task-relevant mnemonic representations can be maintained in the early visual and high-level parietal and frontal cortices, suggesting the highly selective nature of VWM representations.

Material and methods

Participants

Eleven participants from Dartmouth College (19–31 years of age, 2 males) participated in the experiment. All had normal or corrected-to-normal vision and normal color vision, and were neurologically intact. Participants provided informed consent in accordance with the Institutional Review Board of Dartmouth College before the experiment and were monetarily compensated for their participation. All participants were naive to the purpose of the study.

Stimuli and procedure

Stimuli were generated and presented using Matlab and Psychtoolbox 3 (Brainard, 1997; Pelli, 1997). Participants performed a color or orientation delay-estimation task (Wilken and Ma, 2004) in the scanner while fixating at the central dot. Each trial began with the presentation of two square-wave gratings (radius = 5°, spatial frequency = 2 cycles/°, phase angle randomized between 0° and 180°), one in each hemifield (eccentricity = 7°). The orientations of the gratings were randomly chosen from eight orientations that are equally spaced between 0° and 180° in a step of 22.5°. The colors of the gratings were randomly chosen from eight colors that are equally spaced on the CIEL*a*b* color space (L* = 45, a* = 0, b* = 0). All eight colors were isoluminant to the background, determined by a minimum motion task (Cavanagh et al., 1984) prior to scanning of each participant.

Fig. 1. Color and orientation delay-estimation tasks and behavioral performance. A. Participants (N = 11) performed a color or orientation delay-estimation task, each of which was conducted in a separate session. On each trial, they viewed two gratings, one in each hemifield, and were cued to remember either color or orientation of one of the gratings. After a delay of 8.6 s, participants adjusted a bar (clockwise or counterclockwise rotation) until it pointed to the cued color or orientation of the cued grating on the color or orientation wheel in 5 s. Images were enhanced for demonstration purposes. B. Behavioral performance for the color and orientation VWM tasks, modeled by the mixture model (Bays et al., 2009; Zhang and Luck, 2008). Concentration represented the response variability (standard deviation), and pT, pNT, and pU represented the probabilities of responses to target, non-target, and random guessing, respectively. Error bars indicate ± 1 SEM.

The cued and uncued gratings always had different colors and orientations. Each color and orientation was shown four times within a run, twice each in the cued and uncued object in random order. A random jitter of 0–3° were added to each orientation and color (0–3° on a color wheel) on each trial. The two sample gratings were presented for 1 s, followed by a short blank interval of 0.4 s, and then a black arrow at the center of the screen indicating which of the two gratings to remember (left or right). The likelihood of either grating being cued was equal. Participants only needed to remember the color or orientation of the cued grating, depending on the current task requirement. The cue duration was 1 s, followed by a long retention period of 8.6 s. At the end of the delay period, participants were asked to reproduce the cued color or orientation on a color or orientation wheel (radius = 10°). Specifically, they rotated a bar via button presses until it pointed to the cued color or orientation on the wheel as precisely as possible. The angular change brought by each button press (clockwise or counter-clockwise rotation) was 5° for the color task and 2.5° for the orientation task. 5 s was given to complete the adjustment. The color wheel remained identical across trials (Fig. 1A). The memory recall error was measured for each trial as the absolute angular difference (in degrees) between the cued grating and participant’s response. The range of recall errors in the color task was [-π, π], and the range of errors in the orientation task was [-π/2, π/2]. Each run consisted of sixteen trials, which lasted 344 s. Inter-trial-intervals (ITI) varied in 2, 4, 6, or 8 s and there was a 10 s blank fixation period at the beginning of each run. Each participant finished twelve runs each for color and orientation tasks, except one performed ten runs each due to the time limit. Before scanning, each participant practiced the working memory task outside the scanner for 30 min. The color and orientation tasks were performed in separate sessions on different days.

In addition to the main experimental runs, participants also performed several localization runs, including retinotopic mapping, visual field localizer (two runs of 296 s each) and working memory localizer (two runs of 344 s each). The retinotopic mapping session was conducted in a separate session. On each trial, they viewed two gratings, one in each hemifield, and were cued to remember either color or orientation of one of the gratings. After a delay of 8.6 s, participants adjusted a bar (clockwise or counterclockwise rotation) until it pointed to the cued color or orientation of the cued grating on the color or orientation wheel in 5 s. Images were enhanced for demonstration purposes.
conducted following standard retinotopic mapping procedures (Engel et al., 1994; Sereno et al., 1995). The visual field localizer consisted of alternating, blocked presentations of a flickering checkerboard at one of the sample locations (radius = 5°, eccentricity = 7°, block duration = 12 s). Participants fixated at the central fixation and pressed the button when the fixation checker flipped its polarity. The working memory localizer was identical to the main working memory runs except that participants were asked to remember both the color and orientation of the cued grating, and the ITI was fixed at 12 s.

Data acquisition

MRI data were acquired using a Philips Intera Achieva 3T MRI scanner with a 32-channel head coil. Anatomical images were obtained using a standard T1-weighted MPRAGE sequence (voxel size = 1 x 1 x 1 mm). Functional imaging was conducted using a gradient-echo-echo-planar imaging (EPI) sequence (TR = 2 s, TE = 35 ms, FOV = 240 x 240 mm, voxel size = 3 x 3 x 3 mm, flip angle = 90°). 33 slices oriented along the calcarine fissure were acquired to primarily cover the occipital, parietal and frontal cortices.

Mixture-modeling of behavioral responses

A mixture-model (Bays et al., 2009; Zhang and Luck, 2008) was used to assess participants’ behavioral performance. This model is a probabilistic model, which proposes that there are three possible sources of error during the memory task: Gaussian variability in memory for the target, Gaussian variability in memory for the non-target, and random guessing. In our experiment, these three error sources were characterized by the following equation:

\[ p(\theta) = mT \sigma_\theta mNT \sigma_\theta U + \sigma_U / 2\pi \]

where \( \hat{\theta} \) was participant’s response value, \( \theta \) was the target feature value, \( \theta^* \) was the non-target feature value, \( \sigma_\theta \) was a Von Mises distribution with mean of zero and standard deviation \( \sigma \), \( \sigma \) represented the response variability (concentration), and \( mT \), \( mNT \), and \( mU \) represented the probabilities of responses to the target, responses to the non-target, and guessing rate respectively. This model was applied to the color and orientation tasks separately.

Data preprocessing

Functional MRI data were preprocessed using AFNI (http://afni.nimh.nih.gov; Cox, 1996). The data were first registered to the final volume of each scan, and to the retinotopy session. The data were then motion corrected, linearly and quadratically detrended, and z-score normalized within each run. Localizer data were spatially smoothed with a 4-mm FWHM (Full Width at Half Maximum) Gaussian kernel. For the group-level whole-brain analyses, the data were further normalized to the Talairach standard space (Talairach and Tournoux, 1988).

All functional ROIs were defined on the cortical surface reconstructed with FreeSurfer (http://surfer.nmr.mgh.harvard.edu; Fischl et al., 1999, 2001) using the independent functional localizer runs. The visual retinotopic cortex was parcellated into V1, V2, V3, and V4 using standard retinotopic mapping procedures (Engel et al., 1994; Sereno et al., 1995). We only included the voxels in the V1–V4 ROI that showed greater activation during presentation of visual stimuli at each stimulus location (left or right to the fixation) compared to the fixation baseline in the visual field localizer runs (false discovery rate (FDR), \( p < 10^{-4} \)). To define the parietal and frontal ROIs, we first selected the voxels that showed significantly greater activation during the sample period of the working memory localizer task \( (p < 0.01) \) in frontal and parietal regions. The activated voxels were further constrained using anatomical labels from the Destrieux Atlas (Destrieux et al., 2010) provided in FreeSurfer. Specifically, IPS was defined as voxels within intraparietal sulcus (IPS) and the inferior part of superior parietal lobule (SPL), FEF was defined as voxels within superior precentral sulcus (sPCS) and the caudal part of superior frontal gyrus (SFG), and iPCS was defined as voxels within inferior precentral sulcus (iPCS) and the caudal part of inferior frontal sulcus (IFS).

Inverted encoding model

Inverted encoding model analyses were conducted using in-house Python and Matlab scripts, and were performed for color and orientation separately. We extracted population-level feature-selective responses within each ROI using an encoding model of color or orientation introduced by Brouwer and Heeger (2009, 2011), and Ester and colleagues (2013). The feature selectivity of each voxel was characterized as a weighted sum of eight hypothesized channels. The idealized feature tuning curve of each channel is a half-wave-rectified and squared sinusoid raised to the sixth power (FWHM = 0.94 in radians for orientation; FWHM = 1.88 in radians for color) as below:

\[ f(\tau) = \sin(\tau)^6 \]

Broader basis functions raised to the second power (FWHM = 1.57 in radians for orientation; FWHM = 3.14 in radians for color) were used later in the control analysis.

Before feeding the preprocessed data into the inverted encoding model, a baseline from each voxel’s response was removed in each run using the equation: \( B = B - m(m^TB) \) (Brouwer and Heeger, 2011), in which \( B \) represented the data matrix from each run with size \( n \times c \) (n: the number of voxels in the ROI; c: the number of orientations/colors) and \( m \) represented the mean response across all stimulus conditions of length \( n \).

We then computed the weight matrix (\( W \)) that projects the hypothesized channel responses (\( C_k \)) to actual measured fMRI signals in the training dataset (\( B_1 \)) and extracted the estimated channel responses (\( \hat{C}_k \)) for the test dataset (\( B_2 \)) using this weight matrix. The relationship between the training dataset (\( B_1 \), \( n \times n \): the number of repeated measurements) and the channel responses (\( C_k \), \( k \times n \)) was characterized by:

\[ B_1 = W C_1 \]

where \( W \) was the weight matrix (\( n \times k \)).

Therefore, the least-squared estimate of the weight matrix (\( \hat{W} \)) was calculated using linear regression:

\[ \hat{W} = B_1 C_1^T (C_1 C_1^T)^{-1} \]

The channel responses (\( \hat{C}_k \)) for the test dataset (\( B_2 \)) was then estimated using the weight matrix (\( \hat{W} \)):

\[ \hat{C}_k = (\hat{W}^T \hat{W})^{-1} \hat{W}^T B_2 \]

In order to minimize possible contaminations from the sample or probe display, we defined the time period from 8 to 14 s following each trial onset as the retention period. The time series of each voxel in each ROI during this period was extracted, and the average signals across these time points were used as input for the inverted encoding model. Next, we used a leave-one-run-out procedure to build the weight matrix and to calculate the estimated channel outputs for each of the eight orientations or colors in the test dataset. The weight matrix of the encoding model was calculated based on all trials in the training dataset and applied to the test dataset. The estimated channel outputs obtained after each iteration were shifted to a common center, where \( \theta^* \) represented the cued orientation or color channel on each trial. The shifted channel outputs were then averaged across all iterations.
Whole-brain analysis

In addition to the ROIs defined within each participant, we also performed a group-level whole-brain analysis to localize other brain regions that showed greater activation during delay compared to a fixation baseline ($p < 0.03$). Three separate boxcar regressors were used to localize brain areas that were significantly activated during memory encoding, maintenance, and retrieval. Only the voxels activated during maintenance were included in this analysis.

Pattern classification

To compare the results from the inverted encoding model with those from the decoding method, we applied linear Support Vector Machines (SVMs) implemented within the PyMVPA (Hanke et al., 2009) toolbox to the preprocessed fMRI data. Similar to the inverted encoding model analyses, the data from only time points that corresponded to the memory retention period (8–14 s after stimuli onset) were selected on each trial and averaged within each feature in each run. We then used a leave-one-run-out cross-validation procedure to perform an 8-way classification on the spatial patterns of voxel responses for different colors or orientations. The analysis was conducted for the color and orientation tasks separately.

Permutation significance test

Since the hypothesized channels in the inverted encoding model were not independent, we performed permutation tests to determine statistical significance. For each ROI, we randomly shuffled the trial labels for the entire dataset, generated corresponding model channel outputs, and repeated this procedure 10,000 times. Next, we calculated the statistics (e.g., memory index in Results) on each output and obtained a null distribution of 10,000 values. The statistics for a specific comparison was compared with this null distribution and the probability of obtaining the original statistics given this distribution was reported as the $p$-value of the permutation test. A similar permutation procedure was also conducted for the classification analysis.

Results

Population-level feature reconstructions during VWM maintenance in the occipital, parietal, and frontal cortices

Participants ($N = 11$) performed either a color or orientation delay-estimation task in the scanner. Their behavioral responses were modeled with a mixture-model (Bays et al., 2009; Zhang and Luck, 2008) separately for the color and orientation tasks. The average response variability of the color and orientation tasks was $16.71^\circ$ ($\pm 2.96^\circ$) and $15.58^\circ$ ($\pm 1.91^\circ$), and the probability of target responses was $91.5\%$ ($\pm 5.5\%$) and $80.6\%$ ($\pm 5.8\%$) respectively, suggesting that participants maintained high-fidelity representations of color and orientation in respective tasks (Fig. 1B, also see Supplementary Fig 1 for distributions of response errors in each participant).

In order to examine the feature-specific representations in VWM maintenance, we built an inverted encoding model to extract population-level feature-selective response profiles for color and orientation separately over the delay period (Brouwer and Heeger, 2009, 2011; Ester et al., 2013), from ROIs in visual as well as higher-order parietal and frontal areas that have been previously shown to be involved in VWM maintenance (Harrison and Tong, 2009; Offen et al., 2010; Serences et al., 2009; Srimland and Curtis, 2008; Todd and Marois, 2004; Xu and Chun, 2006; see Fig. 2A for the ROIs from a representative participant). This model assumes that the response of each voxel can be characterized by the weighted sum of hypothesized feature channels. By computing the weights that relate the voxel responses to the feature channel responses in the training dataset and applying these weights to the test dataset (a leave-one-run-out procedure), we obtained the estimated responses for each feature channel in each ROI. Since on each trial two objects were displayed (one was cued and the other was not) and each object consisted of two features (one was task-relevant and the other was not), we obtained four sets of reconstructed feature representations: representation for the task-relevant feature of the cued object, the task-irrelevant feature of the cued object, the task-relevant feature of the uncued object, and the task-irrelevant feature of the uncued object. The four feature representations were obtained separately by feeding corresponding feature labels into the inverted encoding model.

Previous studies showed that mnemonic representations in the visual retinotopic cortex were not only found in the region of the cortex contralateral to the cued object, but also in the region ipsilateral to it, suggesting that working memory representations are spatially global (Ester et al., 2009, 2015). Thus, in subsequent analyses, regardless of the retinotopic locations of the cued object, we used all trials to train and test the inverted encoding model in each lateralized ROI, and averaged the results from the corresponding left and right ROIs as the global feature reconstruction result for a specific ROI.

As shown in Figs. 3A and 4A, our results revealed robust population-level feature reconstructions during delay in the occipital, parietal, and frontal cortices, which peaked at the color or orientation of the cued object. The memory-related feature representations were shown for both orientation and color, suggesting that subregions in the parietal and frontal cortices can not only maintain the spatial feature, such as orientation (Ester et al., 2015), but also represent the non-spatial surface feature, such as color. In contrast, reconstructed feature responses for the task-irrelevant feature of the cued object and for the features of the uncued object were significantly reduced or almost absent, suggesting the highly selective nature of the memory-related representations in the retinotopic cortex as well as high-level frontal and parietal regions.

The strength of representations for feature value during VWM maintenance

In order to quantitatively compare the reconstructed feature representations across tasks and objects, we employed several new indices. We first labeled the channel that corresponds to the target feature and the two channels that are closest to it as “Target Channels” (e.g., $-45^\circ$, $0^\circ$, and $45^\circ$ for color and $-22.5^\circ$, $0^\circ$, and $22.5^\circ$ for orientation), and the three channels that are farthest from the target channel as “Non-target Channels”. We then defined the difference between the average response to the Target Channels and that to the Non-target Channels as the Representation Strength Index (RSI). The RSI was computed separately for each feature representation, and was used as input to all subsequent analyses. An RSI reflects the quality of the reconstructed feature representations; a positive RSI would indicate significant feature representations in each condition and the higher the RSI, the stronger the feature representations.

Using the RSI, we first examined the strength of the mnemonic representations for feature value in each ROI. Specifically, we compared the memory-related feature reconstructions for the task-relevant feature of the cued object (e.g., cued color in the color task) with those for the task-relevant feature of the uncued object (e.g., uncued color in the color task). As both features were encoded before the cue appeared, this comparison allowed us to examine the strength of the reconstructed feature representations that were specific to the cued feature value during delay. We defined a Between-object Memory Modulation Index (Between-object MMI), which was computed as the difference between the RSI for the task-relevant feature of the cued object and that of the uncued object. The Between-object MMI was calculated for color and orientation separately and the two Between-object MMIs were averaged to test the robustness of memory for feature value. A
positive Between-object MMI would indicate better feature reconstructions for the cued feature value, compared with the uncued feature value, reflecting specific memory for the cued feature value.

Our results on the Between-object MMI in each ROI showed that robust memory-related representations that were specific for the cued feature value was maintained in IPS \((p = 0.007)\) and FEF \((p = 0.001)\), as well as in the visual retinotopic cortex \((V1–V3; \text{all } p < 0.00001; \text{V4v: } p = 0.007)\). In contrast, no such representations were found in iPCS \((p = 0.166)\) (Fig. 3B). All the \(p\)-values in this and subsequent analyses were corrected for multiple comparisons across ROIs using the FDR method.

**The strength of task-modulated representations during VWM maintenance**

Next, by using a Within-object Memory Modulation Index (MMI), we examined if the representations for features within a cued object were modulated by task. The Within-object MMI characterizes the difference between the reconstructed representations for the feature when it was cued and task-relevant (e.g., color of the cued object in the color task) and those for the same feature of the cued object when it was task-irrelevant (e.g., color of the cued object in the orientation task). Specifically, the Within-object MMI was computed as the difference between the RSIs for the feature of the cued object when it was task-relevant and when it was task-irrelevant. Again, we computed the Within-object MMI for color and orientation tasks separately, and averaged them to examine the robustness of the task modulation effect. A positive Within-object MMI would indicate significant task modulation of the reconstructed feature representations within the same object.

Our analysis on the Within-object MMI in each ROI revealed a significant task modulation effect in the occipital cortex \((V1–V3; \text{all } p < 0.00001)\) and a marginally significant effect in V4v \((p = 0.057)\). This result is consistent with the previous study using a decoding technique that showed successful classification of the cued feature in the early visual cortex \((\text{Serences et al., 2009})\). Importantly, our study also showed that the memory-related feature representations in subregions in the parietal and frontal cortices, such as IPS \((p = 0.0004)\), FEF \((p = 0.015)\), and iPCS \((p = 0.015)\), were significantly modulated by task (Fig. 4B).

We replicated our findings using an earlier time window during delay \((6–12 \text{ s after trial onset})\). The results were very similar to \(8–14 \text{ s}\), with significant Between- and Within-object MMIs in most of the ROIs (Between-object MMI: \(V1: p = 0.008; V2: p < 0.00001; V3: p < 0.00001; V4v: p = 0.012; \text{IPS: } p = 0.0002; \text{iPCS: } p = 0.012\); Within-object MMI: \(V1: p = 0.016; V2: p = 0.007; V3: p = 0.005; V4v: p = 0.055; \text{IPS: } p = 0.038; \text{FEF: } p = 0.013; \text{iPCS: } p = 0.029; \text{Supplementary Fig 2}\)).

We also performed a classification analysis using a linear SVM to examine if the results from encoding and decoding methods were comparable. The Between- and Within-object MMIs were obtained by replacing RSIs with classification accuracies. Again, significant Between- and Within-object MMIs were observed in most of the ROIs (Between-object MMI: \(V1–V3: \text{all } p < 0.00001\); V4v: \(p = 0.0003\); \text{IPS: } p = 0.0003; \text{iPCS: } p = 0.025); Within-object MMI: \(V1–V3: \text{all } p < 0.00001\); V4v: \(p = 0.029; \text{IPS: } p = 0.099; \text{iPCS: } p = 0.007)\), but not in FEF (Between-object MMI: \(p = 0.145); \text{Within-object MMI: } p = 0.197; \text{Supplementary Fig 3}\), indicating that the results from the encoding and decoding approaches support the same conclusions.

**The robustness of individual feature representations**

In the previous analyses, we demonstrated a strong task modulation effect within the same object, as well as robust representations for the cued feature value only. As a last step, we examined the robustness of individual feature representations by examining the significance of their RSIs (Fig. 5). Our results revealed significant representations for the cued color in the color task in the visual retinotopic cortex \((V1–V4v: \text{all } p < 0.00001)\), and also in IPS \((p = 0.0004)\) and iPCS \((p = 0.036)\). No such feature reconstructions were found in FEF \((p = 0.121)\). We also observed significant feature representations for the cued orientation in the orientation task in V1–V3 \((p < 0.00001)\), IPS \((p = 0.002)\), and FEF \((p = 0.0004)\), but not in V4v \((p = 0.366)\) or iPCS \((p = 0.366)\). This result reveals that color- and orientation-selective mnemonic representations are maintained over delay in a distributed neural network, encompassing from the visual retinotopic cortex to higher-order frontoparietal regions.

Second, we examined if the task-irrelevant feature of the cued object was also maintained during the delay period. Previous studies on object-based attention suggested that when one feature of an object is attended, processing of other features within the same object is also facilitated \((\text{Duncan, 1984}; \text{Egly et al., 1994}; \text{O’Craven et al., 1999})\).
Contrary to this prediction, our results showed that representations for the task-irrelevant feature of the cued object (e.g., the orientation of the cued object when the color task was performed or the color of the cued object when the orientation task was performed) were largely abolished. None of the ROIs showed significant feature reconstructions for the task-irrelevant feature of the cued object (all $p > 0.243$).

Third, we examined if there were significant feature reconstructions for the feature of the uncued object when that feature dimension was task-relevant. Specifically, we examined if there were significant feature representations for the color of the uncued object in the color task and
the orientation of the uncued object in the orientation task. Again, none of the feature reconstructions were statistically significant in any of the ROIs we examined (all \( p > 0.265 \)). Furthermore, the task-irrelevant features of the uncued object (i.e., color of the uncued object in the orientation task and orientation of the uncued object in the color task) were not maintained in most of the ROIs (all \( p > 0.142 \)). The only exceptions were found for uncued orientation in the color task in V2 and IPS (both \( p = 0.038 \)).

We performed additional analyses to validate the reliability of the feature representations. First, we repeated the same analysis separately
on retinotopic visual ROIs that were contralateral or ipsilateral to the target locations. This analysis yielded qualitatively similar results to global feature reconstructions that were computed by averaging the results from the corresponding left and right ROIs, as reported in the previous text (Supplementary Fig 4). Only the task-relevant feature in the cued object was maintained in V1–V4v for color (contralateral: V1–V3: all \( p < 0.00001; \) V4v: \( p = 0.038 \); ipsilateral: V1: \( p < 0.00001; \) V2: \( p = 0.0004; \) V3: \( p = 0.0008; \) V4v: \( p = 0.0005 \)), and in V1–V3 for orientation (contralateral: V1: \( p = 0.021; \) V2: \( p = 0.0002; \) V3: \( p < 0.00001; \) ipsilateral: V1: \( p = 0.008; \) V2: \( p < 0.00001; \) V3: \( p < 0.00001 \)), in both the contralateral and ipsilateral ROIs. None of the other features were maintained (all \( p > 0.227 \)).

Next, we collapsed the left and right ROIs of each brain region and built the encoding model on bilateral ROIs to test if the results remained the same. Our results again showed that representation for the cued color in the color task was significant in V1–V4v (all \( p < 0.00001 \)), IPS (\( p = 0.003 \)), and marginally significant in iPCS (\( p = 0.053 \)); and representation for the cued orientation in the orientation task was significant in V1–V3 (all \( p < 0.00001 \)), IPS (\( p = 0.016 \)), and FEF (\( p = 0.006 \)). None of the other task-irrelevant and uncued features were maintained (all \( p > 0.109 \)) (Supplementary Fig 5).

Although we demonstrated the robustness of the representations for the task-relevant feature of the cued object only, there remains a possibility that weaker but significant feature representation was still present for the task-irrelevant feature of the cued object, or task-relevant features of the uncued object. For example, those task-irrelevant or uncued features may be represented, though more coarsely, compared with the task-relevant feature of the cued object (Swan et al., 2016; Shin and Ma, 2016). To test this hypothesis, we reconstructed the feature representations using basis functions with a broader width (see Material and methods for details), and repeated the analysis for the RSI of each feature representation. Changing the width of the basis functions did not significantly affect our results (all \( p > 0.135 \) except for uncued orientation in the color task in V2 and IPS, both \( p = 0.039 \), Supplementary Fig 6).

Another possibility is that the absence of representation for the task-irrelevant and uncued features was due to the differences in model weights extracted in each condition, since different trials or labels were used to build the encoding model in each condition. To exclude this possibility, we reconstructed the feature representations for the task-relevant, uncued conditions and task-irrelevant, cued conditions using the same model weights computed from the task-relevant, cued conditions. Specifically, the model weights for color were calculated from the cued color in the color task, and were applied to both the uncued color in the color task and colors in the orientation task. Likewise the model weights for orientation were calculated from the cued orientation in the orientation task, and were applied to the uncued orientation in the orientation task and orientations in the color task. All the estimations were performed using a leave-one-run-out procedure. Again, our results showed that none of the RSIs were significant (all \( p > 0.125 \), except for the uncued color in the orientation task in V1 (\( p = 0.039 \)), and a trend in V2 and FEF (both \( p = 0.098 \), Supplementary Fig 7), suggesting that neither the task-relevant feature of the uncued object nor the task-irrelevant feature of the cued object was retained.

Taken together, these results confirmed that only the task-relevant feature of the cued object was robustly and consistently maintained during the retention period.

Whole-brain analysis

Lastly, we performed a group-level, whole-brain analysis to localize other brain regions that showed significant task-modulated feature reconstructions during delay. We observed significantly higher BOLD signals during delay in subregions in the parietal and frontal cortices (See Table 1 for a complete list of group-level ROIs and Fig. 2B for the activation map of the group-level analysis). We also observed significant task-modulated responses in the parietal cortex (e.g., inferior parietal lobule) and the posterior frontal cortex (e.g., precentral gyrus), but not in the anterior parts of the frontal cortex (e.g., middle and inferior frontal gyri). This result, together with the results from the previous analyses using the ROIs defined in each individual participant, demonstrated converging evidence that significant task-modulated feature representations are maintained in the occipital, parietal, and frontal cortices during VWM.
regions reported in previous studies are due to unevenly distributed parietal and frontal cortices, this stimulations for non-spatial features (i.e., colors) were also robust in the et al., 2007; Schluppeck et al., 2005). In our study, since representa-

Another source of signals could be saccadic eye movement (Kastner orientation-selective responses in the parietal and frontal areas. This di-

participants remembered the orientation of an object, their attention (Sprague and Serences, 2013; Sprague et al., 2014). Thus, when regions while a visuospatial working memory task was performed spatial-tuning topographic maps were also shown in the frontoparietal memory delay were elicited by non-memory-related sources of signals fi

Table 1

<table>
<thead>
<tr>
<th>ROI Name</th>
<th>Talairach Coordinates [L,P,S]</th>
<th>ROI size (voxels)</th>
<th>p-value (Within-object MMI)</th>
<th>p-value (Between-object MMI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Superior Parietal Lobule</td>
<td>−24.6, −62.2, 47.3</td>
<td>313, 162</td>
<td>0.121</td>
<td>0.520</td>
</tr>
<tr>
<td>Inferior Parietal Lobule</td>
<td>−38.4, −48.2, 42.4</td>
<td>210, 104</td>
<td>0.043*</td>
<td>0.521</td>
</tr>
<tr>
<td>Postcentral Sulcus</td>
<td>41.4, −34.7, 40.6</td>
<td>137, 139</td>
<td>0.077*</td>
<td>0.130</td>
</tr>
<tr>
<td>Postcentral Gyrus</td>
<td>−43.3, −29.1, 45.9</td>
<td>167, 216</td>
<td>0.0001***</td>
<td>0.020*</td>
</tr>
<tr>
<td>Central Sulcus</td>
<td>−37.0, −27.4, 51.2</td>
<td>116, 151</td>
<td>0.045*</td>
<td>0.325</td>
</tr>
<tr>
<td>Precentral Gyrus</td>
<td>−35.6, −13.5, 54.3</td>
<td>79, 210</td>
<td>0.009**</td>
<td>0.314</td>
</tr>
<tr>
<td>Superior Precentral Sulcus (Brodmann area 6)</td>
<td>−24.6, −70.5, 54</td>
<td>182, 202</td>
<td>0.151</td>
<td>0.748</td>
</tr>
<tr>
<td>Inferior Precentral Sulcus (Brodmann area 9)</td>
<td>−43.5, 1.7, 31.3</td>
<td>138, 148</td>
<td>0.342</td>
<td>0.569</td>
</tr>
<tr>
<td>Inferior Frontal Gyrus</td>
<td>−45.2, 15.5, 24.2</td>
<td>246, 77</td>
<td>0.149</td>
<td>0.054*</td>
</tr>
<tr>
<td>Middle Frontal Gyrus</td>
<td>−41.8, 24.9, 29.9</td>
<td>233, 138</td>
<td>1.000</td>
<td>0.606</td>
</tr>
<tr>
<td>Medial Frontal Gyrus</td>
<td>−4.0, 7.7, 49.1</td>
<td>232, 127</td>
<td>0.0005***</td>
<td>0.240</td>
</tr>
<tr>
<td>Insula</td>
<td>−29.0, 21.3, 10.5</td>
<td>91, 139</td>
<td>0.007**</td>
<td>0.008**</td>
</tr>
<tr>
<td>Occipitotemporal Cortex</td>
<td>−52.0, −48.6, −7.8</td>
<td>142, 87</td>
<td>0.021*</td>
<td>0.099*</td>
</tr>
<tr>
<td>Cingulate Gyrus</td>
<td>−3.8, 9.2, 28.8</td>
<td>59, 267</td>
<td>0.210</td>
<td>0.079*</td>
</tr>
</tbody>
</table>

*p < 0.10, *p < 0.05, **p < 0.01. lh: left hemisphere; rh: right hemisphere. All p-values remained uncorrected.

Discussion

In the current study, we investigated how the VWM representation of each feature in multi-feature objects was modulated by task. Using an inverted encoding model, we demonstrated highly task-selective VWM representations in the occipital, parietal (e.g., IPS), and frontal (e.g. FEF and iPCS) cortices. Robust feature-selective representations were reconstructed for the task-relevant feature of the cued object only, and no significant representations were found for the task-invariant feature of the cued object, or for any features of the uncued object. This result suggests that the feature representations are specific to the task-relevant feature only, and are not likely to be elicited by attention to the entire cued object or to the entire task-relevant feature dimension.

Furthermore, these task-modulated neural codes in the parietal and frontal cortices were not limited to a feature that has spatial structures (e.g., orientation), but were also observed with other non-spatial surface features (e.g., color). Previous studies showing orientation-selective responses in the parietal and frontal cortices during VWM maintenance suggest that the frontoparietal regions can maintain feature-specific information over delay (Albers et al., 2013; Bettencourt and Xu, 2016; Ester et al., 2015). However, there is a possibility that these orientation-specific responses observed during memory delay were elicited by non-memory-related sources of signals in the parietal and frontal cortices. For instance, it has been established that the parietal and frontal cortices are involved in the control of spatial attention in a topographic manner (Colby and Goldberg, 1999; Kastner et al., 2007; Silver et al., 2005; Swisher et al., 2007). Robust spatial-tuning topographic maps were also shown in the frontoparietal regions while a visuospatial working memory task was performed (Sprague and Serences, 2013; Sprague et al., 2014). Thus, when participants remembered the orientation of an object, their attention might be deployed along the oriented lines. This differentially distributed spatial attention over each orientation could lead to seemingly orientation-selective responses in the parietal and frontal areas. Another source of signals could be saccadic eye movement (Kastner et al., 2007; Schluppeck et al., 2005). In our study, since representations for non-spatial features (i.e., colors) were also robust in the parietal and frontal cortices, this finding excluded the possibility that feature-specific, memory-related representations in the frontoparietal regions reported in previous studies are due to unevenly distributed spatial attention, motor intention, or eye movements along the oriented lines.

We also localized several anterior regions (e.g. inferior and middle frontal gyrus) of the frontal cortex that were significantly activated during delay, using either individual or group-level analysis. However, the task-modulated feature representations were significantly reduced, compared with those in the posterior frontal regions (e.g., FEF and precentral gyrus). It is consistent with previous findings on the gradient of the neural codes in the prefrontal cortex, which suggest that the representations in the prefrontal cortex are hierarchical, and that more abstract representations are maintained in the more anterior part of the frontal cortex (Badre and D’Esposito, 2007). The maintenance of visual details in our study may thus not require task-modulated representations in anterior frontal regions. These results suggest that the task-modulated neural responses during VWM may be constrained by the functional specialization of the specific regions of the cortex. Relatedly, although a significant task modulation effect was observed both in FEF and iPCS, significant representations for the cued feature value were only found for the orientation task but not for the color task in FEF, and vice versa in iPCS. A similar pattern to iPCS was also observed in V4v. The representation of one feature but not the other in some brain areas suggests that different feature memory involves partially overlapping (e.g., V1-V3, IPS) yet distinct (e.g., V4v, FEF, iPCS) brain networks.

Our results can also provide new insights into long-standing questions in behavioral studies on VWM. In the VWM literature, there have been many efforts to measure the memory precision of task-relevant features. However, due to the methodological limitations, it remains less clear whether the task-irrelevant features of the attended objects were also remembered. When addressing this issue using various behavioral paradigms with either indirect or direct approaches, previous research has reported mixed findings (Marshall and Bays, 2013; Swan et al., 2016; Shin and Ma, 2016; Woodman and Vogel, 2008). In our experiment, we demonstrated that only the task-relevant feature of the cued multi-feature object was maintained in the population-level feature-selective representations, regardless of whether color or orientation was remembered. Similar findings were also reported in studies showing no evidence of active neural representations for unattended items in VWM (LaRocque et al., 2016; Lewis-Peacock et al., 2012). Although some behavioral studies demonstrated robust encoding of the task-irrelevant features (e.g., Marshall and Bays, 2013), the discrepancy between the result in our study and these previous studies could be due to the difference in the length of the memory delays (Xu, 2010), or the difference in encoding strategy. It may be easier for participants to maintain a highly selective representation for the task-relevant feature when they only need to remember one feature dimension throughout a whole session as in the current study, compared to when they have to switch between different feature dimensions within a single session. Also, note that it was still possible that the task-irrelevant, or uncued features were represented at a
subthreshold level (Mendoza-Halliday et al., 2014), which could be difficult for our current method to detect. Our results show a hint of weak representations of the task-irrelevant feature in the uncued object in some of the analyses. However, as these representations were not robust, the reliability of the signals remains to be further investigated. Moreover, since a behavioral measure of participants' memory performance on the task-irrelevant features was not obtained in this experiment, there lacks direct behavioral evidence of whether those task-irrelevant features were remembered during the VWM task. Nevertheless, our results clearly showed that task-relevant and task-irrelevant features are not represented in the same way, which has been implicated in some previous behavioral studies (Swan et al., 2016; Shin and Ma, 2016). Collectively, these results suggest a highly selective mechanism of VWM that protects mnemonic representations from multiple sources of distractions.

One remaining question is where these selective mnemonic signals arise. Recent VWM research has enabled us to rethink the roles of the parietal and frontal regions in VWM, and suggests that higher-order parietal and frontal cortices are not only the site for controlling the memory process, but also potential locations for memory storage. In the current study, we further showed that the mnemonic representations in the occipital, parietal, and frontal cortices are task-modulated. On one hand, these representations are distinct from the task-related signals in previous studies (Courtney et al., 1997; Riggall and Postle, 2012; Warden and Miller, 2010), as they not only reflect the specific types of tasks participants were performing, but also the information content remembered during different tasks. On the other hand, our study lacks a direct measure of the source of the feature representations, and it is thus not clear whether the feature-selective signals are generated from high-level cortical areas per se, or are diffused from activities in the low-level sensory areas. Furthermore, since the topography of stimulus representations in the early visual cortex are different from those in the higher cortical areas (Riley and Constantinidis, 2016), it is more likely that there exists fundamental differences between the task-modulated mnemonic representations in the early visual cortex and those in the parietal and frontal cortices. For example, a recent study has suggested that mnemonic representation in the parietal cortex is more resistant to visual distraction, compared with that in the early visual cortex (Bettencourt and Xu, 2016). It remains to be examined in future research how the memory-related, population-level, feature-selective responses revealed in our study and previous studies (Ester et al., 2013, 2015) differ between the early visual cortex and higher-order frontoparietal regions, and what roles respective regions play in representing and maintaining mnemonic information. This distinction may help us better understand the mixed findings in the current literature on the nature of feature-specific information in the parietal and frontal cortices during VWM maintenance (Albers et al., 2013; Bettencourt and Xu, 2016; Ester et al., 2015; LaRocque et al., 2016; Riggall and Postle, 2012).

To summarize, in the current study, we investigated the task relevance of the feature-specific mnemonic representations in the occipital, parietal and frontal cortices. Our results demonstrated that both spatial and non-spatial features of a remembered object can be maintained in the occipital, parietal and frontal cortices while a VWM task was performed, which extended our understanding of the memory-related content-specific neural codes in the parietal and frontal cortices. Our results also suggest the highly selective nature of such neural codes, by showing that only the task-relevant feature of the remembered multi-feature object was represented. This evidence helps to resolve the long-standing debate on the status of task-irrelevant features during VWM maintenance.

Author contributions
Q.Y. and W.M.S. designed the experiment. Q.Y. conducted the experiment and analyzed the data. W.M.S. supervised the entire project. Q.Y. and W.M.S. wrote the manuscript.

Acknowledgements

Funding: This work was supported by the Burke Award and IBS-R015-D1 to W.M.S. We would like to thank the reviewers for their helpful comments on the previous versions of the manuscript.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.neuroimage.2017.05.055.

References


O’Craven, K., Downing, P., Kanwisher, N., 1999. fMRI evidence for objects as the units of attentional selection. Nature 401, 584–587.


