Title:

Different states of priority recruit different neural codes in visual working memory

Authors:

Qing Yu[1,*] and Bradley R. Postle[1,2,*]

Affiliations:

[1]Department of Psychiatry, University of Wisconsin-Madison, Madison, WI 53719, USA

[2]Department of Psychology, University of Wisconsin-Madison, Madison, WI 53706, USA

***Correspondence should be addressed to:**

Qing Yu

Department of Psychiatry

University of Wisconsin-Madison

Madison, WI 53719, USA

Email: qyu55@wisc.edu

Or:

Bradley R. Postle

Department of Psychology

University of Wisconsin-Madison

Madison, WI 53706, USA

Email: postle@wisc.edu

Abstract

We tracked the neural representation of information with different priority ("attended memory items, AMI" and "unattended memory items, UMI"), using multivariate inverted encoding models with fMRI data from different stages of multiple tasks. Although representation of the identity of AMI and of the UMI was found in a broad brain network, including early visual, parietal and frontal cortex, the identity of the UMI was actively represented in early visual cortex in a distinct "reversed" code, suggesting early visual cortex as a site of the focus of attention. The location context of the AMI and of the UMI was also broadly represented, although only frontoparietal regions supported the simultaneous, priority-tagged representation of the location of all items in working memory. Our results suggest that a dynamic interplay between multiplexed stimulus representations and a frontoparietal salience map may underlie the flexible control of behavior.

Introduction

Important for understanding the flexible control of behavior[1,2] is understanding working memory, the mental retention of task-relevant information and the ability to manipulate it and use this information to guide contextually appropriate actions[3,4]. State-based theoretical models of working memory posit that information can be held at different levels of priority in working memory, with information at the highest level of priority in the focus of attention (FoA), and the remaining information in a variously named state of "activated long-term memory"[5] or "region of direct access"[6]. Much of the empirical support for these models comes from tasks using a "retrocuing" procedure in

which, after a trial's to-be-remembered information has been removed from view, a subset of that information is cued to indicate that it will be tested. Retrocuing can both improve memory performance behaviorally[7] and increase the strength of retrocued information neutrally[8].

The retrocuing procedure allows for the controlled study of the back-and-forth switching of priority between memory items that is required for many complicated working memory tasks, such as the n-back[9] and working memory span[10] tasks. In the dual serial retrocuing (DSR) task, two items are initially presented as memoranda, followed by a retrocue that designates one the "attended memory item" (AMI) that will be interrogated by the impending probe. The uncued item cannot be dropped from working memory, however, because following the initial memory probe, a second retrocue may indicate (with $p = 0.5$) that this initially uncued item will be tested by the second memory probe. Thus, following the initial retrocue, the uncued item becomes an "unattended memory item" (UMI)[11]. fMRI and EEG studies of the DSR task have demonstrated that an active representation was only observed for the AMI, but not for the UMI, using multivariate pattern classification (MVPA)[12-14]. Thus, an elevated level of activation, particularly in temporo-occipital networks associated with visual perception, may be a neural correlate of the FoA. The neural bases of the UMI, however, are less clear.

Most DSR studies to date have failed to find MVPA evidence for an active representation of the UMI[12-14], although such a trace can be transiently reactivated with a pulse of transcranial magnetic stimulation (TMS)[15]. The one study that has found evidence for active representations of the UMI localized them to parietal and frontal cortex, in an analysis of fMRI data from 87 subjects[16]. Thus, the current preponderance

of extant data suggests that the neural representation of the UMI may be at a level of sustained activity that is so low as to be at or below the boundary of what can be detected with current methods and conventional set sizes. Although there are mechanisms other than elevated activity that could represent information in working memory[17,18], the work presented here was designed to assess two alternative hypotheses about the neural representation of the UMI that have received less attention to date. One is that the representation of the UMI may be active, but in a representational format fundamentally different from those of AMI, and therefore difficult to detect with MVPA methods. The second is that what may be most prominently maintained in working memory is a representation of the trial-unique context in which the UMI was presented, rather than a representation of stimulus identity per se.

Although MVPA is a powerful analytic technique that can provide evidence of whether two kinds of information are different, it is inherently limited in that it doesn't directly provide information about how they differ. Therefore, in the current study we used multivariate inverted encoding modeling (IEM)[19-22] to evaluate item-level mnemonic representations of AMIs and UMIs. By specifying an explicit model of how stimulus properties are represented in large populations of voxels, we could assess quantitative and qualitative changes in stimulus representation as a function of changes in priority status. IEM may also be a more sensitive method for tracking working memory representations[22].

Our results revealed two important properties of UMI representations: first, rather than being just a "weak AMI", the UMI is actively represented in early visual cortex, in a format that is different from the AMI; second, contextual information about the UMI is

represented differently than information inherent to the stimulus. That is, frontoparietal circuits maintain a representation of the location of both memory items that also encodes their priority status, a property absent from spatial representations in early visual cortex.

Results

Experiment 1

Behavioral results

Participants performed two DSR tasks (*Retain1* and *Retain2*) in the scanner. In the *Retain1* task, although two orientation patches were initially presented as targets, the same one was always cued twice, meaning that the initially cued orientation remained in the focus of attention (i.e., the AMI) for the remainder of the trial, and the uncued item could be dropped from memory ("dropped memory item," DMI). In the *Retain2* task, the initially uncued item became a UMI, because it was possible that it would be cued by the second retrocue (Figure 1a). Accuracy in the *Retain1* (63.9% ± 1.7%) and *Retain2* (67.0% ± 2.0%) tasks did not differ ($t(7) = 1.402$, $p = 0.204$), nor did accuracies for the Stay (67.3% ± 2.0%) and Switch (62.6% ± 2.8%) conditions of the *Retain2* task ($t(7) = 1.856$, $p = 0.106$).

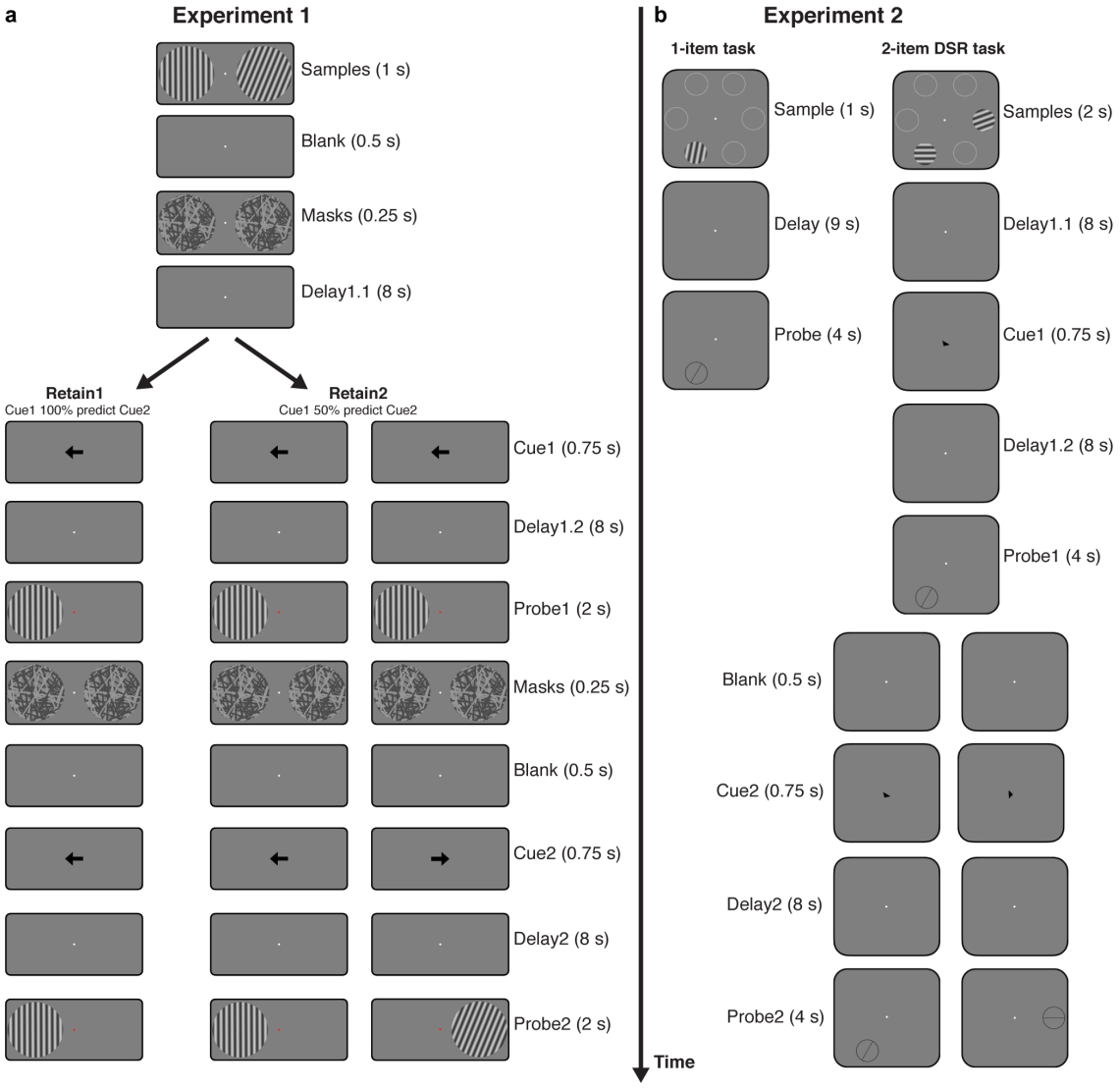

Figure 1. Experimental procedure.

a. In Experiment 1, participants performed two tasks in the scanner in separate blocks. In *Retain1* task, participants remembered two orientations for Delay1.1, one in each hemifield, and were cued on one of them for Delay1.2. After a first probe, the same cue appeared and participants needed to recall the same orientation once again after Delay2. The probe task was a change detection task. In *Retain2* task, participants underwent the same procedure, except that the second cue may switch to the other orientation on 50% of the trials. **b**. In Experiment 2, participants performed the *Retain2* task only, and the two

119 orientations could appear in two of six different locations (white circles are for
120 demonstration purposes and were not present during the actual experiment). Participants
121 performed a delay-estimation task of orientations. Besides the main experiment task,
122 participants also performed a one-item working memory task for training independent
123 IEMs.
124
125 *Reconstructing the neural representation of orientation of the AMI, the DMI, and the*
126 *UMI*
127 Our analytic strategy was to compare IEM reconstructions from models trained on
128 different trial conditions to assess the similarity of representational format between the
129 trained and tested conditions. For Delay1.1, all trials were used to train the IEM, because
130 both items had equal priority status. For Delay1.2, different IEMs were trained using
131 either the AMI or UMI/DMI labels. For AMI-trained IEMs, only the cued stimuli were
132 used to train the IEM, and the IEM was tested on data from both AMI-labeled and
133 UMI/DMI-labeled data. When tested with UMI/DMI-labeled data, reconstructions from
134 this AMI-trained IEM would index the extent to which the representational format of the
135 UMI/DMI was similar to that of the AMI. For UMI/DMI-trained IEMs, the IEM was
136 trained and tested on the uncued stimulus. This IEM allowed us to examine the
137 UMI/DMI representation without being biased by attended information.
138 For Experiment 1 we focused on a Sample-evoked ROI constrained to early visual
139 cortex, because most studies have found robust evidence for an active representation of
140 the AMI in this brain region. Furthermore, no studies, including the Christophel et al.
141 study[16], have found evidence for an active representation of the UMI in this region.

During Delay1.1 (6-8 s after trial onset), when participants had no knowledge of which item in the memory set would be cued, the IEM reconstruction of both was robust, in both the *Retain1* ($p = 0.010$ and $p < 0.00001$) and *Retain2* ($p = 0.031$ and $p < 0.00001$) conditions (Figure 2; two participants were excluded from further analyses due to lack of robust orientation reconstructions in this delay period). Moreover, no significant difference was observed between the two orientation representations in either condition ($ps = 0.431$ and 0.271). All the p-values were corrected across conditions using False Discovery Rate (FDR) method in this and subsequent analyses.

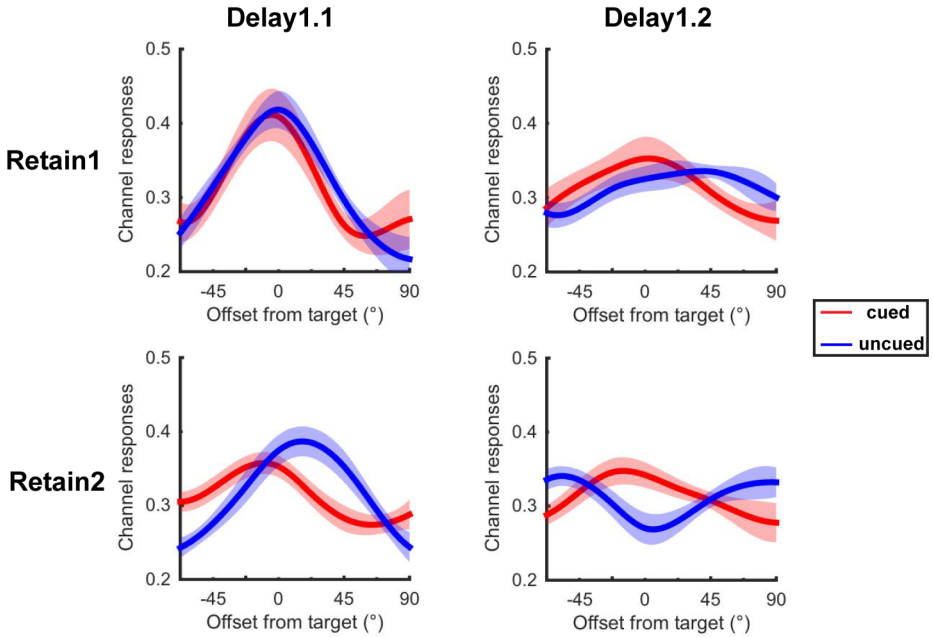

Figure 2. Experiment 1: Orientation reconstruction during different epochs of Delay1 in the Sample-defined visual ROI.

Orientation reconstructions in the *Retain1* and *Retain2* conditions in Delay1.1 (6-8 s) and Delay1.2 (16-18 s). Red line represents the cued orientation (AMI during Delay1.2), and blue line represents the uncued orientation (UMI during Delay1.2). Reconstructions were averaged across all participants. Continuous curves were created with spline interpolation

157 method for demonstration purposes. Channel responses are estimated BOLD responses in
158 relative amplitude. Shaded areas indicate ± 1 SEM.
159
160 For Delay1.2 (the portion of Delay1 that followed the retrocue) we focused on 16-
161 18 s after trial onset (i.e., 6-8 s after retrocue) for maximization of the retrocuing effect.
162 In the *Retain1* condition, robust representation of stimulus orientation was observed for
163 the AMI ($p = 0.037$). In contrast, reconstruction of the DMI was unsuccessful, whether
164 tested with the AMI-trained or the UMI/DMI-trained IEM (ps = 0.424 and 0.915). In the
165 *Retain2* condition, with the AMI-trained IEM, reconstructions of the orientation of the
166 AMI and of the UMI went in opposite directions: a marginally significant positive
167 reconstruction for the AMI ($p = 0.061$) and a significantly negative reconstruction for the
168 UMI ($p = 0.037$). The negative reconstruction of the UMI had the lowest response in the
169 target channel, and progressively higher responses in non-target channels that grew with
170 the distance of the non-target channel increased (Figure 2). The UMI could not be
171 reconstructed with a UMI/DMI-trained IEM ($p = 0.587$; Supplementary Figure 1).
172 The finding of a reliable negative reconstruction for the UMI during late Delay1.2
173 was noteworthy because it deviated from the expectation that we would replicate
174 previous failures to find evidence for an active representation of the UMI during
175 Delay1.2[12-15], It was also inconsistent with the most intuitive alternative account for these
176 previous null findings, which has been that the post-cue representation of the UMI may
177 be qualitatively the same as it was prior to the cue, but the magnitude of its activation has
178 decreased to a level that is no longer detectable. This is because a significant negative
179 reconstruction would require a distributed pattern of activity that differs both from the

trained pattern and from baseline, implying an active representation with a code that is different from, in this case, the code with which the AMI was represented during Delay1.2. Furthermore, this finding would implicate early visual cortex in the active representation of the UMI, which is at variance with accounts positing a privileged role for higher-level regions in visual working memory storage during conditions involving shifting attention[16] or distraction[23,24]. Finally, this finding would represent, to our knowledge, the first report of a negative IEM reconstruction as an interpretable index of the state of an *active* neural representation of stimulus information.

For the reasons listed above, we took several steps to explore possible artifactual explanations for this result. Primarily, we considered the possibility that the negative reconstruction of the orientation of the UMI may have reflected influences from the AMI, because the two could never take the same value on the same trial, but instead always had a distance of at least 22.5°. The reasoning behind this alternative account is that recentering all UMI reconstructions on a common target channel would necessarily produce a situation in which every AMI fell on a non-target channel, and this could result in a negative-going reconstruction after averaging across trials. One reason to doubt this alternative account a priori is because a negative reconstruction was not observed for the DMI in the *Retain1* condition, despite the fact that its procedural conditions were identical. Nonetheless, to assess this possibility analytically, we sorted trials by the distance between the UMI and AMI into four bins (22.5°, 45°, 67.5°, 90°), and obtained reconstructions for these four bins separately. We found that a negative reconstruction of the UMI was obtained for each bin, demonstrating the robustness of a negative UMI reconstruction regardless of the angular distance to the AMI (Supplementary Figure 2).

Furthermore, if the AMI had an influence on UMI reconstruction due to the minimum distance between the two, one would also expect negative reconstruction when testing data from Delay1.1 using labels of the item that would become the UMI in Delay1.2. With this analysis, however, IEM reconstruction failed (i.e., it was not negative; p = 0.816; Supplementary Figure 3).

As an additional step to assess the robustness of the negative reconstruction of the UMI in late Delay1.2, we repeated the analysis using trials from the *Retain2* condition only, to exclude any potential influence from the *Retain1* trials. This analysis, although carried out with only part of the data of the original analysis (50%-67%, depending on the participant), produced a similar negative reconstruction of the UMI (p = 0.049) with an AMI-trained model, and no significant reconstruction of the UMI (p = 0.577) with a UMI-trained model (Supplementary Figure 4).

Experiment 2

Due to its novel and unexpected nature, it was important that we replicate evidence from Experiment 1 for an *active* but *negative* representation of the UMI in early visual cortex. With Experiment 2, we also sought to extend this finding in important ways. First, we would extend our analyses into parietal and frontal regions that have also been implicated in the working memory representation of information. Second, we would investigate in greater detail the representational bases of the UMI by training IEMs with data from a variety of cognitive conditions. Finally, we would investigate whether the representation of an item's trial-specific context might be differently sensitive to changing priority. To elaborate, in Experiment 1 any given orientation patch was presented on one of two locations over the course of an experimental session. This means

that success on any individual trial required not just a memory that a particular item (say, a patch with an orientation of 30°) had been presented at the beginning of the trial, but also a memory of *where* that item had been presented. We have hypothesized that, because maintaining the binding between an item's identity and its context is necessary to keep it in working memory[25,26], this contextual information may be represented in a parietal salience map[27]. Therefore, we designed Experiment 2 to also assess the mnemonic representation of location context by modifying the DSR to feature 6 possible locations at which the two orientation patches could be presented on any trial.

Behavioral results

Experiment 2 required recall responses, which were fit with a 3-factor mixture model (see Methods). The concentration parameter, which estimates the precision of responses, was marginally higher in the Stay condition (16.93 ± 2.74) compared to the Switch condition (11.35 ± 1.67), $t(9) = 2.211$, $p = 0.054$. No such differences were found for any other parameters (probabilities of responses to target: 79.9% ± 1.9% vs. 76.3% ± 3.1%; probabilities of responses to non-target: 3.7% ± 1.7% vs. 4.9% ± 2.4%; probabilities of guessing: 16.4% ± 1.9% vs. 18.8% ± 2.9%), ts < 1.199, ps > 0.261.

Reconstructing representations of the orientation of the AMI and UMI

Besides the AMI- and UMI-trained IEMs as used in Experiment 1, we also trained IEMs on an independent 1-item delayed recall task in Experiment 2, for two reasons: First, these IEMs provided "idealized" estimates of how the brain represents these stimulus properties when only a single stimulus is being processed, thereby excluding any factors that may be associated with processing two stimuli simultaneously; second,

249 independent models were needed to directly compare IEM reconstructions between
250 conditions. *P*-values reported in this section were corrected across conditions and time
251 points within each ROI.

252 *AMI- and UMI-trained IEMs*

253 We first repeated the analyses from Experiment 1, with the difference that the
254 analyses were performed regardless of the retinotopic locations of the stimuli, in order to
255 maximize the number of trials available for each condition. We also applied the IEM
256 analysis to each time point in Delay1.2 to examine how the neural codes changed
257 dynamically with time. In early visual cortex (V1 and V2), patterns of reconstructions of
258 orientation were broadly similar to the findings from Experiment 1 (Figure 3a): AMI-
259 trained IEMs produced significantly positive reconstruction of the AMI in late Delay1.2
260 (ps = 0.002 and 0.036), and significantly negative reconstruction of the UMI (ps = 0.003
261 and 0.036); and UMI-trained IEMs failed to reconstruct the UMIs (ps = 0.654 and 0.475).
262 In IPS, however, we observed a qualitatively different pattern (Figure 3a): robust positive
263 reconstructions of the AMI in all subregions (all ps < 0.049 except in IPS1: p = 0.062)
264 were accompanied by a positive reconstruction of the UMI in IPS5 (ps = 0.019) towards
265 the end of delay; and by positive-trending reconstructions of the UMI in IPS0-2 and IPS4
266 (all ps < 0.098). Also at variance with early visual ROIs, with UMI-trained IEMs the
267 UMI could be successfully reconstructed in IPS5 (p = 0.024), and with positive trends in
268 IPS1 and IPS2 (ps = 0.076 and 0.086). In FEF, the reconstruction of orientation was only
269 successful for the UMI with the AMI-trained IEM, from 14 to 16 s (ps = 0.057 and 0.011;
270 Figure 4a-b). Together, these results indicate that although the UMI could be
271 reconstructed in both early visual cortex (replicating Experiment 1) and in IPS and FEF,

it is represented in a different format in these two regions – different from the AMI in early visual cortex, similar to the AMI in IPS.

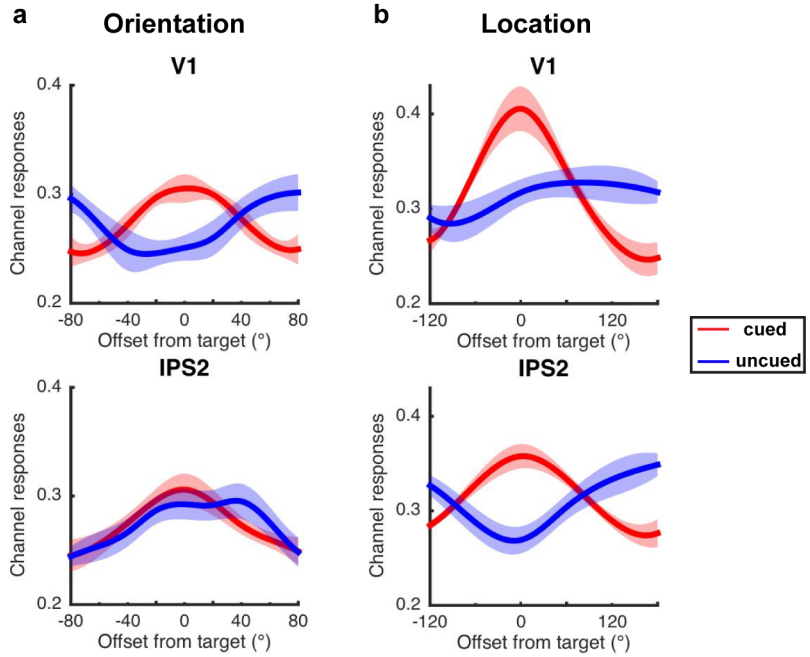

Figure 3. Experiment 2: Orientation and location reconstructions in V1 and IPS2. Demonstration of orientation (**a**) and location (**b**) reconstructions at 18 s in two representative ROIs: V1 (early visual cortex) and IPS2 (parietal cortex), using an AMI-trained model. Red line represents the cued orientation (AMI during Delay1.2), and blue line represents the uncued orientation (UMI during Delay1.2). Reconstructions were averaged across all participants. Continuous curves were created with spline interpolation method for demonstration purposes. Channel responses are estimated BOLD responses in relative amplitude. Shaded areas indicate ± 1 SEM.

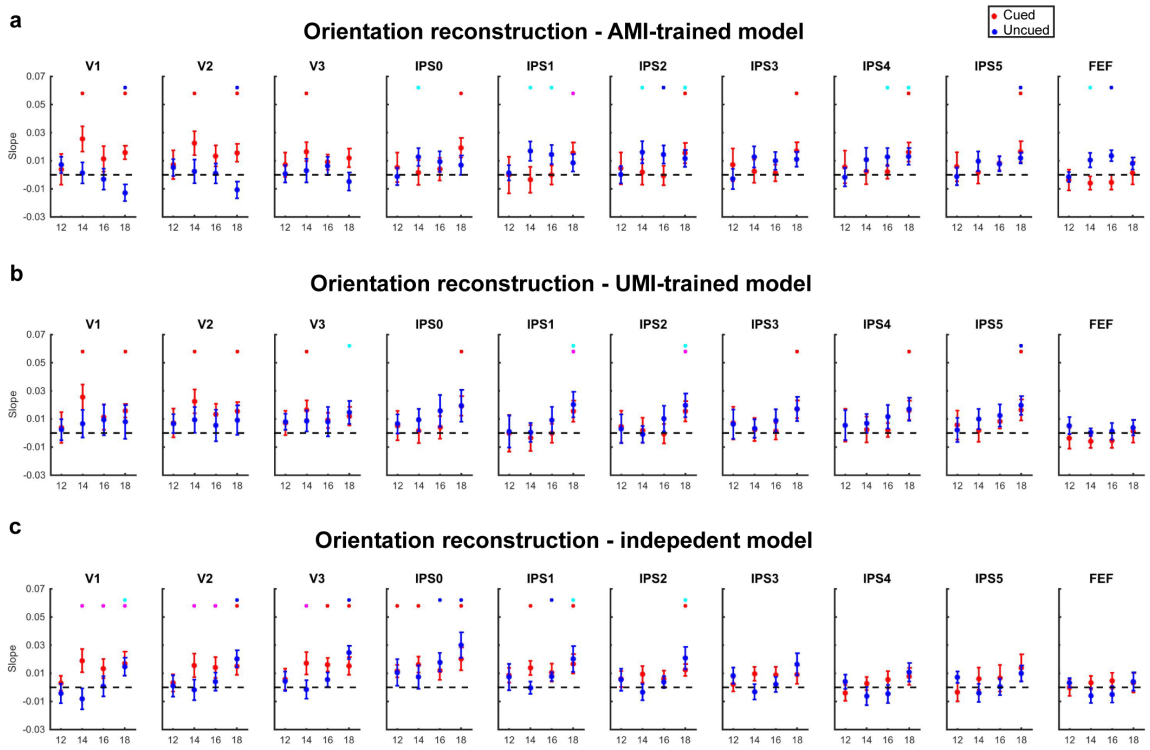

Figure 4. Strength of orientation reconstructions in Delay1.2 in Experiment 2. Slope changes as a function of time during Delay1.2 (12, 14, 16, 18 s after trial onset) for cued (AMI) and uncued (UMI) orientations. Red dots represent the AMI and blue dots represent the UMI. Asterisks at the top of each figure denote the significance of each reconstruction: red asterisk (AMI $p < 0.05$), blue asterisk (UMI $p < 0.05$), magenta asterisk (AMI $p < 0.10$), cyan asterisk (UMI $p < 0.10$). Error bars indicate ± 1 SEM. **a**. Slopes of orientation reconstructions from the AMI-trained IEM. **b**. Slopes of orientation reconstructions from the UMI-trained IEM (red dots are from the AMI-trained IEM for comparison purposes). **C**. Slopes of orientation reconstructions from the independent IEM.

Independent IEMs

297 Next, we sought to reconstruct the orientations of the AMI and UMI using models
298 trained with data from the independent 1-item delayed-recall task. For reconstructions of
299 stimulus orientation we used an IEM trained with data from the TR beginning 4 s after
300 sample onset. In the early visual cortex ROIs (V1-V3) reconstructions of the AMI started
301 to emerge after 14 s and sustained across Delay1.2 ($p = 0.087$ in V1 and ps < 0.036 in V2
302 and V3 at 18 s). For the UMI in the same ROIs, in contrast, reconstructions of the UMI
303 were not significant across the initial 6 s of Delay1.2, before becoming positive for the
304 final TR before probe onset ($p = 0.087$ in V1 and ps < 0.00001 in V2 and V3). In the
305 caudal IPS (IPS0-2), in contrast to early visual regions, reconstructions of the AMI and of
306 the UMI with the independent IEM both followed a similar pattern of steadily
307 strengthening across the delay period (all ps < 0.018 except for UMI in IPS2 ($p = 0.060$)
308 and in IPS3 ($p = 0.050$) at 18 s). No reconstructions were successful in rostral IPS ROIs,
309 nor in FEF (Figure 4c).
310
311 *Reconstructing representations of the location of the AMI and UMI*
312 *AMI- and UMI-trained IEMs*
313 In early visual cortex, whereas the location of the AMI could be reconstructed
314 across the entirety of Delay1.2 with an AMI-trained IEM (all ps < 0.026, except for one
315 time point (14 s) in V1 ($p = 0.213$) and in V2 ($p = 0.207$)), the location of UMI could
316 only be reconstructed during one early TR (14 s), all ps < 0.048 (Figure 3b). In IPS and
317 FEF, in contrast, although there was some variability across ROIs, the general pattern
318 was of positive and sustained reconstruction of the location of the AMI (all ps < 0.017 at
319 18 s except $p = 0.074$ in IPS1), and of negative -- and also sustained -- reconstruction of

the location of the UMI (all *p*s < 0.034 at 18 s except *p*s = 0.067 and 0.095 in IPS2 and IPS4, Figure 3b; Figure 5a). This pattern resembled that of orientation reconstruction in early visual cortex.

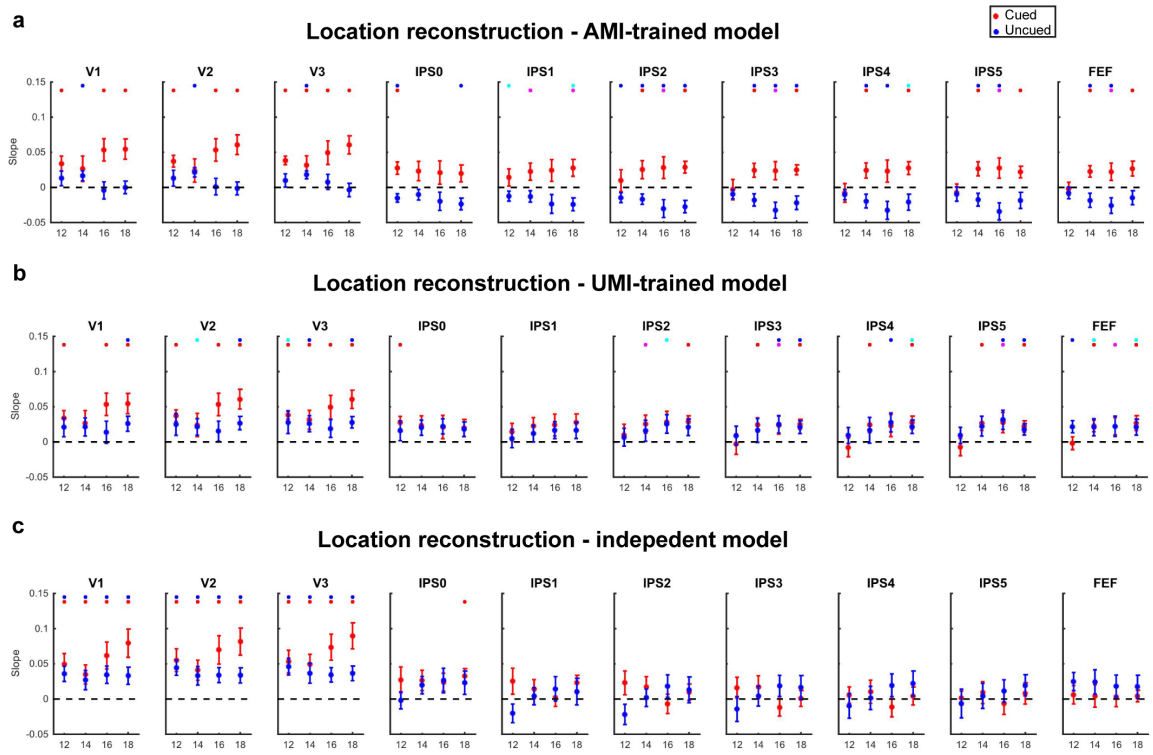

Figure 5. Strength of location reconstructions in Delay1.2 in Experiment 2. Slope changes as a function of time during Delay1.2 (12, 14, 16, 18 s after trial onset) for cued (AMI) and uncued (UMI) locations. Red dots represent the AMI and blue dots represent the UMI. Asterisks at the top of each figure denote the significance of each reconstruction: red asterisk (AMI $p < 0.05$), blue asterisk (UMI $p < 0.05$), magenta asterisk (AMI $p < 0.10$), cyan asterisk (UMI $p < 0.10$). Error bars indicate ± 1 SEM. **a**. Slopes of location reconstructions from the AMI-trained IEM. **b**. Slopes of location reconstructions from the UMI-trained IEM (red dots are from the AMI-trained IEM for comparison purposes). **C**. Slopes of location reconstructions from the independent IEM.

Turning to UMI-trained IEMs, in stark contrast to what was observed for reconstruction of orientation, the location of the UMI could be reconstructed in regions of both early visual cortex and rostral IPS, especially at the late TR (18 s, all *p*s < 0.048 except IPS4: $p = 0.053$). Results in FEF with both AMI- and UMI-trained IEMs mirrored those from rostral IPS (Figure 5b).

Independent IEMs

For reconstructions with an IEM from the independent 1-item task we used a "delay" IEM trained with data from the TR beginning 10 s after sample onset (i.e., the end of delay period). In early visual ROIs, the location of both AMI and UMI could be successfully reconstructed, across all TRs of Delay1.2, with this independent IEM (all *p*s < 0.044). In IPS and in FEF, in contrast, stimulus location could not be reconstructed in any ROI (except for the AMI at 18 s in IPS0, $p = 0.010$; Figure 5c).

Although these analyses were intended to measure the working-memory representation of location context, an alternative account was possible: The successful reconstruction, in early visual cortex, of stimulus location during Delay1.2 may have merely reflected lingering activation patterns from the allocation of external attention to the trial-initiating presentation of sample stimuli. To confirm the interpretability of these results in terms of the working-memory representation of location context, we extended these analyses to Delay2, by which time no stimulus had occupied the retinotopic location of the UMI for 24 s, and Cue2 had updated the status of item to either DMI (on Stay trials) or AMI (on Switch) trials. In early visual ROIs, using the independent IEM, the strength of the representation of the location of the previously unattended item remained significantly positive for the Delay2 (all *p*s < 0.033 except at 32 s in V1, $p =$

0.154) on Switch trials. On Stay trials, in contrast, these reconstructions declined and became null in V1 and V2 (all *p*s > 0.352), and negative at 32 s for V3 ($p = 0.011$; Supplementary Figure 5), suggesting the differentiation between location representations in early visual cortex on Stay and Switch trials.

Discussion

It is commonly accepted that neural representations of information, including of information held in working memory, are supported by anatomically distributed networks. What remains unclear is the extent to in which stimulus-related patterns of activity that can be localized to different brain regions may employ the same or different representational formats, and may support similar or different functions. In the current study we manipulated the momentary state of priority of information in working memory, and employed multivariate encoding models to track interregional differences and dynamic transformations in the representation of behaviorally relevant information.

Dynamic, multiplexed representation of stimulus identity in visual working memory

With regard to the representation of stimulus identity (here, orientation), our results indicate that early visual cortex supports multi-dimensional representation of stimulus identity: the representation of the AMI is maintained relatively stably across the delay period, and the representation of the UMI follows a more dynamic trajectory, and only emerges when memory probe onset is imminent; the two representations share some features in common as both of them can be reconstructed using an independent IEM, but they also differ from each other, manifesting as the negative reconstruction of the UMI relative to the AMI. Although subregions in IPS and FEF also maintain some

representations of the AMI and UMI, the critical difference is a positive-AMI-encoded representation of the UMI, rather than a negative one, is observed in IPS and FEF.

The fact that information with different attentional priority is represented in different neural codes in early visual cortex but not in parietal and frontal regions supports the view that the former is the primary site for the focus of attention in visual working memory, an observation consistent with sensorimotor-recruitment models of visual working memory[4,28]. AMI-encoded representations of the UMI, as well as UMI-encoded representations of the UMI, were identified in several IPS ROIs and in FEF, a pattern consistent with a recent study using multivariate decoding techniques[16]. Additionally, a novel finding from Experiments 1 and 2 was evidence for a reverse-AMI-encoded representation of the UMI in early visual cortex that emerged late in the delay period. Representations in an anatomically distinct network[16,23,29], or in early visual cortex but with one or more codes that are different from a sensory code, could both be effective and mutually compatible schemes for protecting information from interference.

With regard to the time course of stimulus representation across the delay period, the emergence, at the end of Delay1.2, of an AMI-encoded representation of the AMI in the IPS is consistent with the idea that prioritization in working memory initiates a reconfiguration of the representational state of that information in preparation for memory-guided action[7].

Robust and distributed representation of location context

Although our DSR task explicitly tested visual working memory for a nonspatial stimulus feature, the task can nevertheless not be performed successfully without the trial-specific representation of the location at which each stimulus was presented. Indeed,

403 context binding may be essential of working memory[25,26]. Furthermore, many studies
404 have demonstrated the automatic binding of location information to the to-be-
405 remembered visual features[30-33].

406 Because delay-period BOLD signal intensity in IPS is markedly higher on trials
407 that require visual working memory for 3 items drawn from the same category than for 3
408 items drawn from different categories[27], it may be that IPS recruitment scales with
409 demands on context binding. This would be consistent with the idea that a frontoparietal
410 salience map tracks the location context of items held in visual working memory. In
411 Experiment 2, although the location representations of the AMI and of the UMI were
412 robust across the delay period, the patterns were differently sensitive to attentional
413 priority in different brain regions. Whereas early visual regions supported AMI-encoded
414 representations of the location of the AMI but not of the UMI, the pattern in IPS and FEF
415 was different. In addition to supporting AMI-encoded representations of the location of
416 the AMI, IPS and FEF also, and simultaneously, supported reverse-AMI-encoded
417 representations of the location of the UMI. Thus, unlike early visual cortex, this
418 frontoparietal system represented the location of all items in working memory, and the
419 priority status associated with those locations. Qualitatively, this pattern of results is
420 reversed from what was observed for the representation of orientation. This is consistent
421 with the idea that context and priority in visual working memory are represented by the
422 same frontoparietal salience map that tracks these factors during behaviors that do not
423 make any overt demands on working memory[34-36].

424 *Negative reconstructions of the representation of orientation and of location context*

Although our results make clear that many brain areas can simultaneously represent the same information, often in similar representational formats, it seems unlikely that any two region's functions are completely redundant. Rather, we interpret our results as reflecting multiple graded distributions of functional activity, with the likelihood that, for some circuits in some instances, the primary function being supported is one other than storage, per se. The late-in-the-delay emergence of AMI-encoded representations of the AMI in IPS may be one example. Nonetheless, the delay-spanning representation of stimulus information (a.k.a., "storage") is a cardinal property of working memory, and we propose that the recoding of stimulus information into a reverse-AMI-encoded representation may be a mechanism for accomplishing this function for stimuli that are in working memory but outside the focus of attention.

It has been noted that the requirement of temporarily storing information in a noisy neuronal network, for later retrieval, is mathematically equivalent to transmitting that information through a noisy channel[37]. Shannon[38] demonstrated that high-fidelity transmission of information though a noisy channel can be accomplished by recoding the message into a format that takes into account the structure of the noise, then decoding it at the receiving end. One possibility is that the "negative reconstructions" that we have observed, in early visual cortex for the representation of the identity of the UMI, and in IPS and FEF for the representation of the location context of the UMI, reflect a common strategy for maintaining a high-fidelity representation of information while it is held in working memory, but outside the FoA. We note that these instances of negative reconstruction can't be characterized as inhibition, because the effect of inhibition should be to "flatten" a representation. Nor are they likely to be the inhibitory engrams

postulated by Barron and colleagues[39], because whereas the effect of the inhibitory engram would be to minimize representation-related activity, the negative reconstructions that we have described here must be the result of an active reconfiguration of activity in all the voxels feeding into that IEM. Thus, although these reverse-AMI-encoded representations are, indeed, quantitatively negative reconstructions, in functional terms it may be more fitting to characterize them as negative to the code on which the IEM was trained.

Methods

Participants

Ten individuals (5 males, mean age 22.8 ± 3.8 years) participated in Experiment 1. Two were excluded from analysis due to lack of orientation reconstruction in the first memory delay (see Results for details). Another ten individuals (4 males, mean age 23.8 ± 3.5 years) participated in Experiment 2. All were recruited from the University of Wisconsin-Madison community. All had normal or corrected-to-normal vision, were neurologically healthy, and provided written informed consent approved by the University of Wisconsin-Madison Health Sciences Institutional Review Board. All participants were monetarily compensated for their participation.

Stimuli and Procedure

All stimuli were created and presented using Matlab and Psychtoolbox 3 extensions.

Experiment 1. Participants performed two dual serial retrocuing (DSR) tasks (*Retain1* and *Retain2*) in the scanner. During the *Retain1* trials, participants viewed two sinusoidal gratings (radius = 5°, contrast = 0.6, spatial frequency = 0.5 cycles/°, phase angle randomized between 0° and 180°) with different orientations presented simultaneously on the screen (one in each hemifield, eccentricity = 7°) for 1 s. After an interval of 0.5 s, two masks composed of random black and white lines were presented at the stimulus location for 0.25 s, followed by the first delay period. After 8 s ("Delay1.1") a retrocue indicating which grating would be tested at the end of the trial appeared for 0.75 s (Cue1). After an additional 8 s ("Delay1.2"), a probe grating requiring a Y/N recognition response was presented for 0.5 s, followed by a response period of 1.5 s (Probe1). Another two masks that were identical to the first two masks were presented after Probe1 for 0.5 s. 0.5 s later, a second cue that was always identical to the first cue appeared for 0.75 s (Cue2), indicating that participants would be tested on the same grating, followed by a delay of 8 s (Delay2). A second probe grating was presented 0.5 s, and 1.5 s was given to make the second response (Probe2). The task for both probes was to judge whether the orientation of the probe grating was the same as the cued grating, and probes were always presented at the same location as the cued grating. Half of the probes had exactly the same orientation as the cued grating, whereas the other half had an orientation difference between 10° to 20°. Intertrial-interval was either 4 s or 6 s. *Retain2* trials had exactly the same procedure as *Retain1* trials, except that Cue1 did not predict Cue2. Therefore, on half of the trials, Cue2 was identical to Cue1, meaning that the same cued orientation would be probed twice (a "*Retain2-stay*" trial); and on the other half Cue2 was different from Cue1, meaning that Probe2 would probe memory for the target

that had not been tested by Probe1 (a "*Retain2-switch*" trial, Figure 1a). Following our previous work, the item cued by Cue1 was termed the AMI and the item that was not cued by Cue1 in *Retain2* condition was termed the UMI. In addition, the item that was not cued by Cue1 in *Retain1* condition was termed the "dropped memory item" (DMI), because it could be dropped from working memory. The two tasks were conducted in separate blocks, and participants were informed which task they would be performing at the beginning of each block. The two orientations on each trial were randomly selected from a fixed set of eight orientations (0°, 22.5°, 45°, 67.5°, 90°, 112.5°, 135°, 157.5° with a random jitter between 0° and 3°). With the constraint that each of the eight orientations appeared once in both locations during each run, and that the two orientations on any given trial could never be the same. This resulted in a minimum distance of 22.5° between the two orientations on every trial. Each run began with an 8-s blank period, was comprised of 16 trials, and lasted 600 s. Six of the participants performed six runs of the *Retain1* task, one performed seven runs and one performed twelve runs. Seven participants performed twelve runs of the *Retain2* task, and one performed fourteen runs.

Experiment 2. Participants performed two working memory tasks in the scanner. The first task was one-item delayed recall (a.k.a. "delayed estimation") of orientation, intended for training IEMs that would be used to analyze data from this experiment's DSR task. On each trial, one grating (radius = 2°, contrast = 0.6, spatial frequency = 0.5 cycles/°, phase angle randomized between 0° and 180°) was presented on the screen with an eccentricity of 7° and participants were asked to remember its orientation. The location of the grating was chosen from six fixed locations (60° of distance from each other), and the orientation of the grating was chosen from nine orientations (0°, 20°, 40°,

60°, 80°, 100°, 120°, 140°, 160°) with a random jitter between 0° and 3°. The grating appeared on the screen for 1 s, followed by a delay period of 9 s, and then by a response period of 4 s. During the response period, an orientation wheel (2° in radius) was presented at the same location as the sample grating, and participants needed to rotate the needle at the center of the wheel to make it match the remembered orientation as precisely as possible. The inter-trial-interval was fixed at 8 s. Each run consisted of eighteen trials, resulting in a run length of 404 s. Participants performed a total of 24 to 30 runs of the one-item working memory task in two separate scan sessions.

The second task was a two-item DSR task testing delayed recall (a.k.a. "delayed estimation") of orientation patches that could appear in any of six possible locations. On each trial, participants viewed two gratings (parameters identical to those in the first task) presented at two of six fixed locations and were asked to remember both. The two gratings appeared on the screen for 2 s, followed by a first delay period (Delay1.1) of 8 s. After that a cue appeared at the center of the screen for 0.75 s, which was a triangle-shaped arrow that pointed to one of the two sample locations. After another 8 s (Delay1.2), an orientation wheel was presented at the same location as the cued grating, and participants needed to reproduce the cued orientation on the wheel within a 4-s response window. 0.5 s after the first response period, participants saw a second cue, 50% of which would point to the first cued location (Stay), and the other 50% would point to the first uncued location (Switch). After a third 8 s of delay (Delay2), a second orientation wheel was presented at the same location as the second-cued grating, and again participants needed to reproduce the cued orientation on the wheel in 4 s (Figure 1b). The inter-trial-interval was fixed at 8 s. Each run consisted of twelve trials, resulting

in a run length of 536 s. Participants performed 12 runs of this DSR task in one scan session.

In both experiments, electrooculography (EOG) of vertical and horizontal eye movements was recorded while participants performed the tasks in the scanner to ensure central fixation throughout each trial.

Behavioral analysis for Experiment 2

We analyzed behavioral responses with a three-factor mixture model[40] that uses maximum likelihood estimation to generate estimates of 1) the proportion of responses based on a representation of the probed item ("responses to target"); 2) the proportion of responses incorrectly based on a representation of the unprobed item (i.e., "misbinding" or "swap" errors); and 3) the proportion of responses that were guesses not based on either memory item; as well as 4) a "concentration" parameter that estimates the precision of target responses. Conceptually, the concentration parameter is similar to a model-free measure of the precision of responses that is computed as the inverse of the standard deviation of the distribution of responses.

Data acquisition

Whole-brain images were acquired using a 3 Tesla GE MR scanner (Discovery MR750; GE Healthcare) at the Lane Neuroimaging Laboratory at the University of Wisconsin–Madison HealthEmotions Research Institute (Department of Psychiatry). Functional imaging was conducted using a gradient-echo echo-planar sequence (2 s repetition time (TR), 22 ms echo time (TE), 60° flip angle) within a 64 × 64 matrix (42

562 axial slices, 3 mm isotropic). A high-resolution T1 image was also acquired for each
563 session with a fast spoiled gradient-recalled-echo sequence (8.2 ms TR, 3.2 ms TE, 12°
564 flip angle, 176 axial slices, 256 × 256 in- plane, 1.0 mm isotropic).
565
566 *Data preprocessing*
567 Functional MRI data were preprocessed using AFNI (http://afni.nimh.nih.gov) [41].
568 The data were first registered to the final volume of each scan, and then to anatomical
569 images of the first scan session. The data were then motion corrected, detrended, and z-
570 score normalized within each run.
571
572 *ROI definition*
573 Anatomical ROIs were created by extracting masks from the probabilistic atlas of
574 Wang and colleagues[42], and warping them to each subject's structural scan in native
575 space.
576 Analyses in Experiment 1 were carried out in a Sample-defined ROI within a
577 merged V1-V3 ROI. In Experiment 1, we modeled each trial with six boxcar regressors:
578 Sample (1 s), Delay1.1 (8 s), Delay1.2 (8 s), Probe1 (2 s), Delay2 (8 s), and Probe2 (2 s).
579 We focused on voxels with the highest sample-evoked response because these tend to
580 show high decoding accuracy of delay-period signal[14,16,43]. Specifically, we selected the
581 top 1000 voxels that responded maximally during the sample period, within the visual
582 cortex (V1-V3 combined). All the analyses were performed in the contralateral
583 retinotopic ROIs.

584 Analyses in Experiment 2 were carried out in individual atlas-defined ROIs,
585 including early visual cortex (V1-V3), IPS (IPS0-IPS5), and FEF. All the analyses were
586 performed in each ROI merged between the right and left hemispheres.
587
588 *Multivariate inverted encoding modeling*
589 We used inverted encoding models (IEMs) to evaluate the representation of
590 orientation (in Experiments 1 and 2) and of location (in Experiment 2) of the AMI and
591 UMI during different trial epochs. The IEM assumes that the responses of each voxel can
592 be characterized by a small number of hypothesized tuning channels. In Experiment 1 the
593 number of orientation tuning channels was eight, and in Experiment 2 the number of
594 orientation tuning channels was nine and the number of location tuning channels was six.
595 Following previous work[22,44], the idealized feature tuning curve of each channel was
596 defined as a half-wave-rectified and squared sinusoid raised to the sixth power (FWHM =
597 0.94 rad) for orientation in Experiment 1, to the eighth power (FWHM = 0.82 rad) for
598 orientation in Experiment 2, and to the sixth power (FWHM = 1.88 in rad) for location in
599 Experiment 2.
600 Before feeding the preprocessed data into the IEM, a baseline from each voxel's
601 response was removed in each run using the following equation from[19]:
602 $B = B - m(m^T B)$
603 in which B represented the data matrix from each run with size $v \times c$ (v: the number of
604 voxels in the ROI; c: the number of orientations/locations) and m represented the mean
605 response across all stimulus conditions of length v. A constant of 100 was added to B to
606 avoid matrix inversion problems after baseline removal.

We then computed the weight matrix (W) that projects the hypothesized channel responses (C_1) to actual measured fMRI signals in the training dataset (B_1), and extracted the estimated channel responses ($\hat{C}_2$) for the test dataset (B_2) using this weight matrix. The relationship between the training dataset (B_1, $v \times n$, n: the number of repeated measurements) and the channel responses (C_1, $k \times n$) was characterized by:

$$B_1 = WC_1$$

Where W was the weight matrix ($v \times k$).

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

$$\widehat{W} = B_1 C_1^T (C_1 C_1^T)^{-1}$$

The channel responses ($\hat{C}_2$) for the test dataset (B_2) was then estimated using the weight matrix ($\widehat{W}$):

$$\hat{C}_2 = (\widehat{W}^T \widehat{W})^{-1} \widehat{W}^T B_2$$

For Experiment 1, we used a leave-one-run-out procedure to build the weight matrix and to calculate the estimated channel outputs for each of eight orientations in the test dataset. IEMs were constructed with average signals across several time points during an epoch of interest. The obtained weight matrices were applied to the same time points in the test dataset. The estimated channel outputs obtained after each iteration were shifted to a common center, with 0° corresponding to the cued orientation channel. The shifted channel outputs were then averaged across all iterations and all time points of interest within each participant. For Experiment 2, multiple IEMs were trained. First, as with Experiment 1, we used a leave-one-run-out procedure to train IEMs on the AMI from Delay1.2 and on the UMI from Delay1.2, on signals at each time point of interest.

Additionally, we trained "independent" IEMs with data from the one-item delayed-recall task, and tested these IEMs on data from the DSR task. We used the TR 4 s after trial onset to train an orientation IEM, and the TR 10 s after trial onset to train a location IEM. All the IEMs were estimated for orientations and locations separately.

To characterize the strength of each reconstruction, we collapsed over the channel responses on both sides of the cued channel, averaged them, and calculated the slope of each collapsed reconstruction using linear regression. A larger positive slope indicates stronger positive representation, and a larger negative slope indicates stronger negative representation. We used a bootstrapping procedure to characterize the significance of the slopes. For each condition, eight (in Experiment1) or ten (in Experiment 2) orientation/location reconstructions were randomly sampled with replacement from the reconstruction pool of eight (in Experiment1) or ten (in Experiment 2) participants and averaged. This procedure was repeated 10000 times, resulting in 10000 average orientation/location reconstructions for each condition, and correspondingly 10000 slopes. To obtain a two-tailed measure of the p values, the probabilities of obtaining a positive (p_{pos}) or negative (p_{neg}) slope among the 10000 slopes was calculated separately, and the p value of the bootstrapping test was calculated using the following equation:

$$p = 2*\min(p_{pos}, p_{neg})$$

Acknowledgements

This work was supported by National Institutes of Health grant R01MH064498 to B.R.P.

Author Contributions

Q.Y. and B.R.P. designed the experiment. Q.Y. conducted the experiment and analyzed the data. Q.Y. and B.R.P. wrote the manuscript.

Competing Interests statement

The authors declare no competing interests.

References

1 Miller, E. K. & Cohen, J. D. An integrative theory of prefrontal cortex function. *Annu Rev Neurosci* **24**, 167-202, doi:10.1146/annurev.neuro.24.1.167 (2001).

2 Stokes, M. G. *et al.* Dynamic coding for cognitive control in prefrontal cortex. *Neuron* **78**, 364-375, doi:10.1016/j.neuron.2013.01.039 (2013).

3 Baddeley, A. Working memory: looking back and looking forward. *Nat Rev Neurosci* **4**, 829-839, doi:10.1038/nrn1201 (2003).

4 D'Esposito, M. & Postle, B. R. The cognitive neuroscience of working memory. *Annu Rev Psychol* **66**, 115-142, doi:10.1146/annurev-psych-010814-015031 (2015).

5 Cowan, N. *Attention and memory : an integrated framework*. (Oxford University Press, 1995).

6 Oberauer, K. Access to information in working memory: exploring the focus of attention. *J Exp Psychol Learn Mem Cogn* **28**, 411-421 (2002).

7 Myers, N. E., Stokes, M. G. & Nobre, A. C. Prioritizing Information during Working Memory: Beyond Sustained Internal Attention. *Trends Cogn Sci* **21**, 449-461, doi:10.1016/j.tics.2017.03.010 (2017).

8 Sprague, T. C., Ester, E. F. & Serences, J. T. Restoring Latent Visual Working Memory Representations in Human Cortex. *Neuron* **91**, 694-707, doi:10.1016/j.neuron.2016.07.006 (2016).

9 Owen, A. M., McMillan, K. M., Laird, A. R. & Bullmore, E. N-back working memory paradigm: a meta-analysis of normative functional neuroimaging studies. *Hum Brain Mapp* **25**, 46-59, doi:10.1002/hbm.20131 (2005).

10 Conway, A. R. *et al.* Working memory span tasks: A methodological review and user's guide. *Psychon Bull Rev* **12**, 769-786 (2005).

11 Larocque, J. J., Lewis-Peacock, J. A. & Postle, B. R. Multiple neural states of representation in short-term memory? It's a matter of attention. *Front Hum Neurosci* **8**, 5, doi:10.3389/fnhum.2014.00005 (2014).

12 Lewis-Peacock, J. A., Drysdale, A. T., Oberauer, K. & Postle, B. R. Neural Evidence for a Distinction between Short-term Memory and the Focus of Attention. *Journal of Cognitive Neuroscience* **24**, 61-79 (2012).

13 LaRocque, J. J., Lewis-Peacock, J. A., Drysdale, A. T., Oberauer, K. & Postle, B. R. Decoding attended information in short-term memory: an EEG study. *J Cogn Neurosci* **25**, 127-142, doi:10.1162/jocn_a_00305 (2013).

14 LaRocque, J. J., Riggall, A. C., Emrich, S. M. & Postle, B. R. Within-Category Decoding of Information in Different Attentional States in Short-Term Memory. *Cereb Cortex* **27**, 4881-4890, doi:10.1093/cercor/bhw283 (2017).

15 Rose, N. S. *et al.* Reactivation of latent working memories with transcranial magnetic stimulation. *Science* **354**, 1136-1139, doi:10.1126/science.aah7011 (2016).

16 Christophel, T. B., Iamshchinina, P., Yan, C., Allefeld, C. & Haynes, J. D. Cortical specialization for attended versus unattended working memory. *Nat Neurosci*, doi:10.1038/s41593-018-0094-4 (2018).

17 Barak, O. & Tsodyks, M. Working models of working memory. *Curr Opin Neurobiol* **25**, 20-24, doi:10.1016/j.conb.2013.10.008 (2014).

18 Wolff, M. J., Jochim, J., Akyurek, E. G. & Stokes, M. G. Dynamic hidden states underlying working-memory-guided behavior. *Nat Neurosci* **20**, 864-871, doi:10.1038/nn.4546 (2017).

19 Brouwer, G. J. & Heeger, D. J. Cross-orientation suppression in human visual cortex. *J Neurophysiol* **106**, 2108-2119, doi:10.1152/jn.00540.2011 (2011).

20 Brouwer, G. J. & Heeger, D. J. Decoding and reconstructing color from responses in human visual cortex. *J Neurosci* **29**, 13992-14003, doi:10.1523/JNEUROSCI.3577-09.2009 (2009).

21 Sprague, T. C. *et al.* Inverted Encoding Models Assay Population-Level Stimulus Representations, Not Single-Unit Neural Tuning. *eNeuro*, doi:https://doi.org/10.1523/ENEURO.0098-18.2018 (2018).

22 Ester, E. F., Sprague, T. C. & Serences, J. T. Parietal and Frontal Cortex Encode Stimulus-Specific Mnemonic Representations during Visual Working Memory. *Neuron* **87**, 893-905, doi:10.1016/j.neuron.2015.07.013 (2015).

23 Xu, Y. Reevaluating the Sensory Account of Visual Working Memory Storage. *Trends Cogn Sci* **21**, 794-815, doi:10.1016/j.tics.2017.06.013 (2017).

24 Lorenc, E. S., Sreenivasan, K. K., Nee, D. E., Vandenbroucke, A. R. E. & D'Esposito, M. Flexible coding of visual working memory representations during distraction. *J Neurosci*, doi:10.1523/JNEUROSCI.3061-17.2018 (2018).

25 Oberauer, K. & Lin, H. Y. An interference model of visual working memory. *Psychol Rev* **124**, 21-59, doi:10.1037/rev0000044 (2017).

26 Schneegans, S. & Bays, P. M. Neural Architecture for Feature Binding in Visual Working Memory. *J Neurosci* **37**, 3913-3925, doi:10.1523/JNEUROSCI.3493-16.2017 (2017).

27 Gosseries, O. *et al.* Parietal-Occipital Interactions Underlying Control- and Representation-Related Processes in Working Memory for Nonspatial Visual Features. *J Neurosci* **38**, 4357-4366, doi:10.1523/JNEUROSCI.2747-17.2018 (2018).

28 Serences, J. T., Ester, E. F., Vogel, E. K. & Awh, E. Stimulus-specific delay activity in human primary visual cortex. *Psychol Sci* **20**, 207-214, doi:10.1111/j.1467-9280.2009.02276.x (2009).

29 Mendoza-Halliday, D., Torres, S. & Martinez-Trujillo, J. C. Sharp emergence of feature-selective sustained activity along the dorsal visual pathway. *Nat Neurosci* **17**, 1255-1262, doi:10.1038/nn.3785 (2014).

30 Foster, J. J., Bsales, E. M., Jaffe, R. J. & Awh, E. Alpha-Band Activity Reveals Spontaneous Representations of Spatial Position in Visual Working Memory. *Curr Biol* **27**, 3216-3223 e3216, doi:10.1016/j.cub.2017.09.031 (2017).

31 Postle, B. R., Awh, E., Serences, J. T., Sutterer, D. W. & D'Esposito, M. The positional-specificity effect reveals a passive-trace contribution to visual short-term memory. *PLoS One* **8**, e83483, doi:10.1371/journal.pone.0083483 (2013).

32 Rajsic, J. & Wilson, D. E. Asymmetrical access to color and location in visual working memory. *Atten Percept Psychophys* **76**, 1902-1913, doi:10.3758/s13414-014-0723-2 (2014).

33 Sereno, A. B. & Amador, S. C. Attention and memory-related responses of neurons in the lateral intraparietal area during spatial and shape-delayed match-to-sample tasks. *J Neurophysiol* **95**, 1078-1098, doi:10.1152/jn.00431.2005 (2006).

34 Bisley, J. W. & Goldberg, M. E. Attention, intention, and priority in the parietal lobe. *Annu Rev Neurosci* **33**, 1-21, doi:10.1146/annurev-neuro-060909-152823 (2010).

35 Jerde, T. A., Merriam, E. P., Riggall, A. C., Hedges, J. H. & Curtis, C. E. Prioritized maps of space in human frontoparietal cortex. *J Neurosci* **32**, 17382-17390, doi:10.1523/JNEUROSCI.3810-12.2012 (2012).

36 Sprague, T. C. & Serences, J. T. Attention modulates spatial priority maps in the human occipital, parietal and frontal cortices. *Nat Neurosci* **16**, 1879-1887, doi:10.1038/nn.3574 (2013).

37 Koyluoglu, O. O., Pertzov, Y., Manohar, S., Husain, M. & Fiete, I. R. Fundamental bound on the persistence and capacity of short-term memory stored as graded persistent activity. *Elife* **6**, doi:10.7554/eLife.22225 (2017).

38 Shannon, C. E. A mathematical theory of communication. *The Bell System Technical Journal* **27**, 379-423, doi:10.1002/j.1538-7305.1948.tb01338.x (1948).

39 Barron, H. C., Vogels, T. P., Behrens, T. E. & Ramaswami, M. Inhibitory engrams in perception and memory. *Proc Natl Acad Sci U S A* **114**, 6666-6674, doi:10.1073/pnas.1701812114 (2017).

40 Bays, P. M., Catalao, R. F. & Husain, M. The precision of visual working memory is set by allocation of a shared resource. *J Vis* **9**, 7 1-11, doi:10.1167/9.10.7 (2009).

41 Cox, R. W. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput Biomed Res* **29**, 162-173 (1996).

42 Wang, L., Mruczek, R. E. B., Arcaro, M. J. & Kastner, S. Probabilistic maps of visual topography in human cortex. *Cerebral Cortex* **25**, 3911-3931 (2015).

43 Emrich, S. M., Riggall, A. C., Larocque, J. J. & Postle, B. R. Distributed patterns of activity in sensory cortex reflect the precision of multiple items maintained in visual short-term memory. *J Neurosci* **33**, 6516-6523, doi:10.1523/JNEUROSCI.5732-12.2013 (2013).

44 Yu, Q. & Shim, W. M. Occipital, parietal, and frontal cortices selectively maintain task-relevant features of multi-feature objects in visual working

779 memory. *Neuroimage* **157**, 97-107, doi:10.1016/j.neuroimage.2017.05.055
780 (2017).
781