Distinct functional roles of distributed cortical representations for working memory storage

Vivien Chopurian[*,1,2], Simon Weber[2,3], Thomas Christophel[*,1,2]

[1]Department of Psychology, Humboldt-Universität zu Berlin, Berlin, Germany
[2]Bernstein Center for Computational Neuroscience Berlin and Berlin Center for Advanced Neuroimaging, Charité Universitätsmedizin Berlin, corporate member of the Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Berlin, Germany
[3]Research Training Group "Extrospection" and Berlin School of Mind and Brain, Humboldt-Universität zu Berlin, Berlin, Germany

[*]Corresponding authors. E-mail: vivien.chopurian@hu-berlin.de; tbchristophel@gmail.com

Abstract

While distributed cortical areas represent working memory contents, their necessity for memory maintenance has been questioned. To understand whether these regions serve separable functional roles when multiple items are maintained, we examined the effect of visual working memory load on neural information across cortical regions. We show that increasing visual load decreased behavioural recall performance and item-specific mnemonic information in visual cortex but not in anterior regions. Both items, irrespective of their serial position, were represented in visual cortex, whereas sPCS maintained only the most recent item. Our results provide evidence for distinct functional roles of visual cortices, where single items are stored with high fidelity, and anterior cortices, where multiple items are represented using different cortical patterns.

Keywords: fMRI; working memory; capacity; serial order

Working memory capacity is limited. The precision of recall decreases as we increase the number of items to remember [1, 2, 3]. In neuroimaging studies, the load dependent decrease in recall performance is accompanied by a decline in the strength of visual working memory representations [4, 5, 6]. While this evidence demonstrates a link between the ability to decode cortical representations and the capacity to memorize information, it remains unclear whether different cortical regions play distinct roles as memory load increases.

Previous studies highlight a distributed network of sensory, parietal and frontal regions to maintain memorized contents during working memory delay [7, 8, 9, 10]. For single visual items, studies have found correlations between decodable information in visual cortex and recall performance across participants and trials [11, 12, 13]. The neuronal selectivity in visual cortex might enable the maintenance of highly detailed, sensory-like, representations of memorized content [14, 11, 15]. These representations have been found to be susceptible to visual distractors, leading to reduced delay-period decoding if the distractors also affected working memory recall [12, 16, 17]. Furthermore, imaging and single-cell studies demonstrate the maintenance of item-specific representations in parietal and prefrontal regions [18, 19, 20, 9, 21]. In non-human primates, persistent delay activity in the prefrontal cortex directly correlated with behavioural performance [22, 23]. Presumably, anterior regions maintain abstracted, sensory-modality independent mnemonic representations [24, 25, 26], which appear to be protected from interference during visual distraction [9, 21, 17, 12].

However, if multiple low-level visual items have to be encoded and maintained in working memory, their representations are likely to compete for resources in visual cortex [27]. Prior work demonstrates that the extent of cortical overlap of multiple representations predicts the degree to which memory recall is impaired [28]. Therefore, we assume that working memory contents from different sensory modalities should interfere less because these representations share little overlap within sensory regions [29, 30]. The recruitment of abstract anterior representations (e.g., verbal labels, categories) could reduce interference by helping to separate similar sensory representations [31, 32, 33] at the cost of precision. Moreover, multiple items might be represented in anterior regions in a form that minimizes interference between them (e.g., by encoding their serial position [34, 35, 36, 37]).

In this study, we investigated the effect of visual working memory load on these distributed cortical representations. We show, by keeping the overall working memory load constant at two items but varying the sensory modality of each item, that increasing visual working memory load selectively reduced mnemonic information about items present in visual cortex but not in anterior regions. After the sequential presentation of the two memory items and the subsequent delay

period, participants were cued to the target item. This allowed us to track individual items depending on their serial order position during the delay period. While there is no difference in decoding accuracy in visual areas for the first and the second orientation, frontal regions maintained more information about the second item in both load conditions. Our results provide evidence for a division of labor, where visual cortex is representing individual items and anterior regions are essential for the maintenance of multiple items.

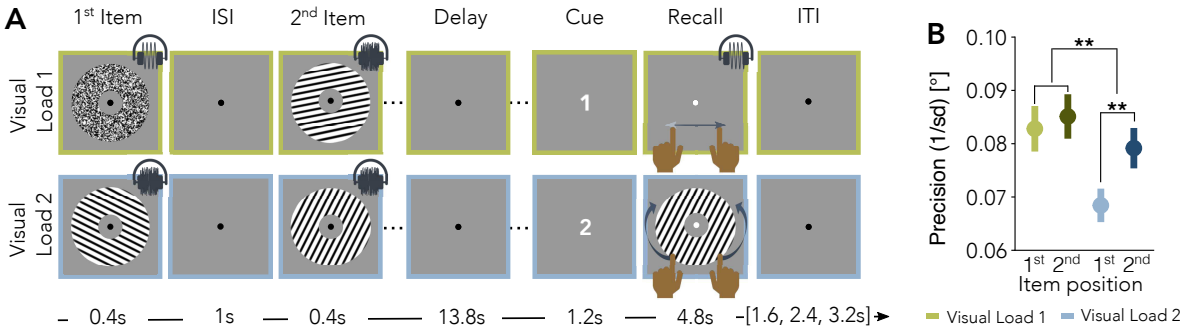


Fig. 1: Task setup and behavioural results. **A** Example experimental trials with varying visual load. The top row shows a visual Load 1 trial, where items from both modalities are shown and the target modality is auditory. The bottom row shows a visual Load 2 trial, where both items are visual. Auditory Load 2 trials (not shown) are included in the design to balance the behavioural task. Visual samples were presented jointly with auditory noise and auditory samples were presented with visual masks. **B** Behavioural results for visual targets. Precision is calculated by taking inverse of the standard deviation (SD) of the deviation from response to target orientation in degrees. Asterisks in indicate significance levels of $*p < .05$, $**p < .01$ or $***p < .001$. Error bars show the SEM.

Results

Visual load reduces mnemonic information in visual cortex

On trials with visual recall, participants performed better on trials where only the target item was an orientation and the other item was a pitch (visual Load 1, $precision = 0.079°$) compared to the visual Load 2 condition (visual Load 2, $precision = 0.070°, t_{(80)} = 2.98, p = 0.004$, Fig.1B). To investigate this visual load effect on the distributed cortical representations, we decoded orientations continuously with a periodic support vector regression (pSVR [13], see Methods and Fig.2A) from fMRI data. While all three regions of interest, the visual cortex (V1-V3), intraparietal sulcus (IPS) and precentral sulcus (sPCS), show information about orientations, the decoding accuracy in each region changed with load and across time (Fig.2B).

Specifically, the decoding accuracy in visual cortex was affected by visual load throughout the trial. Our results showed main effects of load ($F_{(1,80)} = 47.55, p < 0.0001$) and time ($F_{(1,80)} =$

$54.45, p < 0.0001$), as well as an interactions between load and time ($F_{(1,80)} = 14.16, p < 0.001$) on decoding accuracy. Early in the delay (7.2s-11.2s after item onset) and with increased visual working memory load, decoding accuracy decreased ($t_{(80)} = -7.53, p < 0.0001$, Bonferroni corrected). In this early delay phase cortical representations can be assumed to be a mixture of perceptual and mnemonic representations due to the hemodynamic lag. However, the same pattern of results is also present late in the delay (12-16s after item onset), when signals can be attributed to working memory (12s-16s, see [17]; $t_{(80)} = -3.52, p = 0.002$, Bonferroni corrected). While in both conditions decoding accuracy dropped from early to late delay (Load 1: $t_{(80)} = 7.19, p < 0.0001$, Load 2: $t_{(80)} = 4.52, p < 0.0001$,), the decrease in decoding accuracy was stronger for visual Load 1 compared to visual Load 2 ($t_{(80)} = -3.76, p = 0.0003$), but never dropped below chance (Fig.2B).

In contrast, anterior regions were not affected by load late in the delay. In IPS, our analysis revealed main effects of load ($F_{(1,80)} = 7.1, p = 0.009$) and time ($F_{(1,80)} = 7.81, p = 0.006$), as well as an interaction between load and time ($F_{(1,80)} = 5.38, p = 0.23$) on decoding accuracy. Here, decoding accuracy was reduced in for items in visual Load 2 trials compared to Load 1 trials early in the delay ($t_{(80)} = -3.50, p = 0.002$, Bonferroni corrected), but not late in the delay ($t_{(80)} = -0.49, p = 1.0$, Bonferroni corrected). There was a significant difference in the change in decoding accuracy across time between load conditions ($t_{(80)} = -2.32, p = 0.023$). While the decoding accuracy for representations in the visual Load 2 condition showed no difference between the early and late delay period in IPS ($p > 0.6$), the decoding accuracy for the single visual item dropped from early to late delay ($t_{(80)} = 3.29, p = 0.0015$, Bonferroni corrected). In sPCS, there was no difference in decoding accuracy between load conditions in either early or late delay (all $p > 0.6$).

Preference for the most recent item in sPCS

We presented participants with two items sequentially on all trials, which allowed us to examine positional differences in memory recall and neural representations. Considering item position in the behavioural analysis, our results indicated main effects of visual load ($F_{(1,80)} = 11.29, p = 0.001$) and item position ($F_{(1,80)} = 6.74, p = 0.013$) on precision, but no significant interaction of load and item position ($F_{(1,80)} = 3.64, p = 0.06$). When participants memorized two orientations, recall precision was higher for the second orientation ($precision = 0.079°$) compared to the first ($precision = 0.068°; t_{(80)} = -3.53, p = 0.002$, Bonferroni corrected; Fig.1B), consistent with the recency effect [38]. When only the target item was an orientation and the non-target item was a pitch, there was no difference in precision between item positions ($precision_{first} = 0.082°, precision_{second} = 0.085°, t_{(80)} = -0.52, p = 1.0$, Bonferroni corrected).

Next, we analysed whether the distribution of memory representations across cortical regions

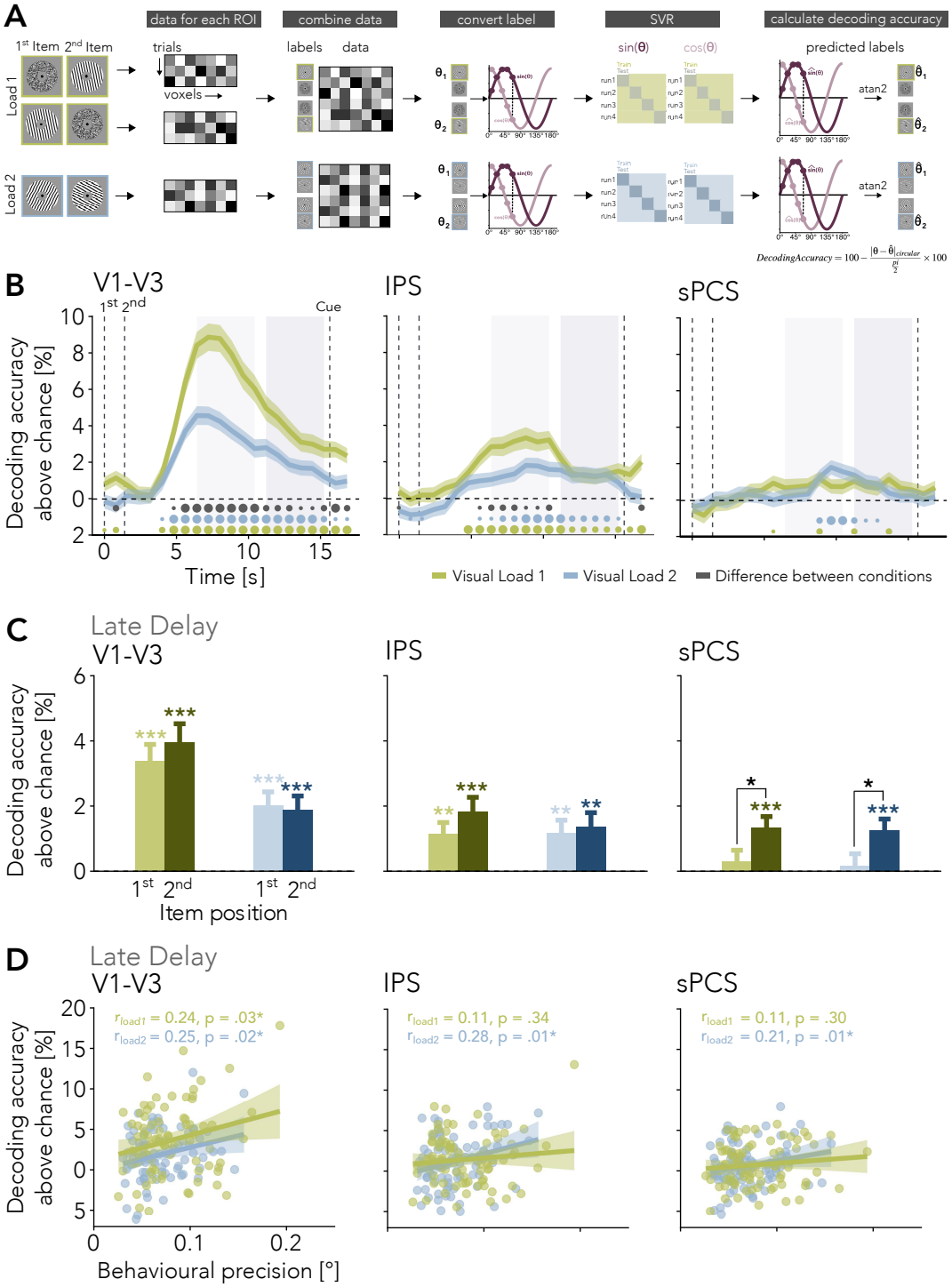


Fig. 2: Analysis steps and decoding results. **A** Analysis steps for each ROI. After selecting each trial for every condition, we combined trials and labels for both target and non-target orientations into one dataset. In order to get linear labels, we converted the orientation labels in degrees into sine and cosine labels. Support vector regression (SVR) was applied to each sine and cosine label [13]. The difference between the labels and predicted labels was converted into decoding accuracy. See Methods for details. **B** Timecourse of item-specific decoding accuracy for each visual Load condition in all three ROIs. Early (7.2s-11.2s) and late delay (12s-16s) after first item onset were chosen to avoid overlap. Decoding accuracies for both items were averaged in each load condition. **C** Decoding accuracy for the first and second visual item in each load condition during the late delay period. Results for the early delay are shown in Supplementary Fig.1. **D** Correlations between late delay decoding accuracy and performance. Points represent single subjects. Dots (small, medium, large) at the bottom of **B** and asterisks in **C** indicate significance levels of $*p < .05$, $*p < .01$ or $***p < .001$. Error bars and shaded ribbons show SEM.

was affected by the sequential presentation order. Early in the delay, we did not find any differences in decoding accuracy for the first and second item (Supplementary Figure 1). Late in the delay, there was no difference between item positions in visual cortex (all $p > 0.5$, Bonferroni corrected) and IPS (all $p > 0.3$, Bonferroni corrected, Fig.2C). However, in sPCS, there was an increase in information for the second item as compared to the first item (main effect order: $F_{(1,80)} = 11.822, p < 0.001$). Notably, this neural recency effect in sPCS was shown in both visual load conditions (Load 1: $t_{(80)} = -2.78, p = 0.02$, Load 2: $t_{(80)} = -2.51, p = 0.041$, Bonferroni corrected) and equal in size (Interaction between item order position and load condition: $F_{(1,80)} = 0.009, p = 0.93$), although the behavioural recency effect was only present on trials where two orientations were presented (see Fig.1B).

Decoding accuracy correlates with behavioural recall precision

Our results so far indicate that anterior as well as posterior regions store working memory representations. To quantify the contributions of the three regions to behavioural performance in each load condition, we correlated decoding accuracy and behavioural precision across all participants. Decodable information about orientations in visual cortex during the late delay correlated with behavioral performance in both visual load conditions ($r_{Load1} = 0.23, p = 0.036, r_{Load2} = 0.23, p = 0.04$). However, behavioural precision correlated with decoding accuracy in IPS ($r = 0.28, p = 0.012$) and sPCS ($r = 0.22, p = 0.048$) only in the Load 2 condition, when visual working memory load was high (Fig.2D). Further analysis comparing the correlations for each ROI showed that they are not significantly different from each other (all $p > 0.20$). A partial correlation revealed a unique contribution of decoding accuracy in visual cortex for behavioural precision in the visual Load 1 condition ($r = 0.25, p = 0.026$), but not with higher visual working memory load.

Shared representations in visual cortex, distinct mnemonic code in sPCS

Considering that we observed changes in decoding accuracy across delay phases, we tested whether the neural code changes across time. First, we probed whether the multi-voxel pattern at one time-point generalizes to other time-points, within each load condition. When a single orientation is maintained in memory, the code at late delay time-points seemed to generalize to other late delay time-points (Fig.3A), indicating a stable mnemonic code in all three regions. However, when increasing visual load, we find significant clusters mainly around the diagonal of the matrix, suggesting subtle changes of mnemonic code over time (Fig.3B).

Next, we tested whether the neural code for orientations is shared between load conditions.

We trained the pSVR on each time-point in one load condition and tested on each time-point in the other load condition. Our findings suggest that these representations within each time-point generalized in visual cortex, where we can decode representations from each load condition from each other load condition (see diagonal in Fig.3 C and D). When training and testing on different time-points (off-diagonal), early perceptual representations did not generalize well to later time-points, likely because of changes in Load 2 representations (see above).

This pattern did not hold up in other regions. In IPS, representations appeared to generalize well during the early stage of trial but diverged from each later in the delay when training and testing on the same time-points (diagonal in Fig.3C and D and Fig.2). We observed asymmetric off-diagonal cross-decoding in IPS, suggesting that representations at earlier time-points in Load 2 were similar to representations at later time-points in Load 1. There was evidence for cross-decoding from Load 1 to Load 2 early in the delay in sPCS, but no reliable cross-decoding late in the delay (see Fig.3C and Supplementary Fig.2). This result indicates distinct mnemonic representations for the two load conditions in sPCS.

Discussion

Visual working memory representations are distributed across the cortex [39, 10]. We investigated the effect of working memory load on these distributed representations by keeping the overall load constant at two items, but varying the sensory modality of each item. Our results suggest that the precision of visual working memory recall is limited by the resources available in visual cortex for the maintenance of multiple similar items. In contrast, increasing visual working memory load did not reduce item-specific information in IPS and sPCS, suggesting that working memory recall relies more on these anterior regions when multiple items are maintained. Information about recent items was increased in sPCS, while the other regions represented the first and second item equally. This indicates that the behavioral recency effect could be a result of such a recency preference of sPCS. In all three cortical regions, the item-specific decoding accuracy predicted behavioural recall precision. Our cross-decoding and temporal generalization analyses suggest that information for two visual items is integrated early in the delay in visual cortex, but stored in distinct mnemonic representations in anterior regions.

Previous studies showed that increasing visual working memory load reduces the ability to decode neural representations of remembered spatial locations and motion directions [4, 6]. While these studies did not selectively investigate representations in visual cortex, the electrodes and voxels included in the respective analyses overlapped with occipital regions. When increasing

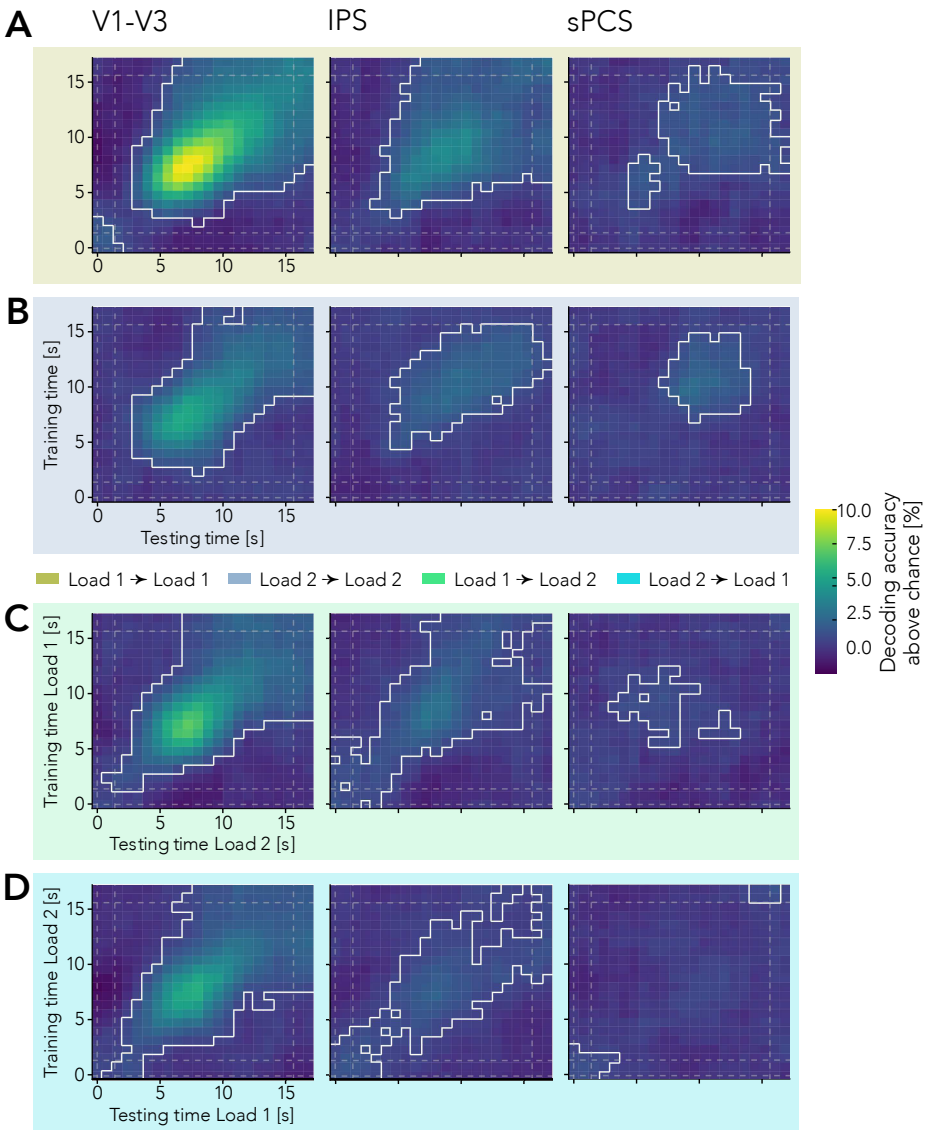


Fig. 3: Temporal generalization and cross-condition decoding results. Temporal generalization results for **A** Visual Load 1 and **B** Visual Load 2 condition. Results for cross-condition temporal generalization are shown in **C**, where we trained on each time-point in Load 1 and tested on each time-point in Load 2 condition and **D**, where we trained on each time-point in Load 2 and tested on each time-point in Load 1. Averages and statistical comparison between each cross decoding in the early and late delay period are shown in Supplementary Figure 2. Dotted lines indicate first item, second item and cue onset. White outlines denote significant clusters (all $p < 0.05$).

visual spatial working memory load, occipital and posterior parietal areas showed a decrease in strength of neural representations, which was also reflected the decrease in behavioural performance [5]. These studies, as well as our task, encouraged the precise recall of the low-level visual features and thus precise neural encoding in early visual cortex. However, the visual cortex does not seem to have the capacity of maintaining more than one of these representations with high fidelity, independent of item order position, resulting in reduced behavioural precision.

Anterior memory representations were not disrupted by increased visual working memory load, consistent with prior studies showing distractor-resistant representations in these regions [21, 16, 17, 12]. In parietal cortex, there was no difference in decoding accuracy between load conditions late in the delay and no difference between item order positions. These representations might support working memory recall for low-level visual features at higher loads. Notably, representations in sPCS selectively encoded the second item late in the delay, even when the first item was from a different sensory modality and no significant behavioural recency effect was found. It is possible that sPCS representations are recruited when visual working memory load is increased and in turn, representations in visual cortex are less reliable than during Load 1. Thus, only at a higher visual load, the neural recency effect is expressed in a behavioural recency effect. The neural recency effect in anterior regions cannot be easily explained by residual perceptual activity or saccades, as it occurs only late in the delay, and was still present in prior work using a non-spatial feature (color) as well as an additional mask after the last item (see [40], Experiment 3). Moreover, this work reports the neural recency effect in sPCS and IPS, suggesting that the division of labor between these cortical regions might be modulated by other task parameters. For example, while in our study the cue was presented after the delay, they retro-cued the target before the delay [40]. Future work is needed to elucidate how serial order and item representations produce neural and behavioural recency effects [41].

Our results suggest a division of labor, where visual cortices are more important for the precise recall of individual low-level sensory items and anterior regions become more important when multiple items need to be maintained in working memory. We demonstrate that different components of visual working memory performance, such as load and recency effects, can be explained by the differential involvement of multiple cortical regions.

1 Methods

Participants

Eighty-six participants were scanned. Four data sets with incomplete recordings (resulting from participant dropouts) and one data set with image artifacts (due to a technical failure of the recording setup) were removed from the analysis. Our final sample consisted of 81 participants (48 female, 4 left handed, $mean_{age} = 25$, $sd_{age} = 4$). All participants provided their informed consent and were compensated with 10€ per hour for their participation. The study was approved by the ethics committee of the Humboldt Universität zu Berlin (id: 2022-40) and conducted according to the principles of the Declaration of Helsinki [42].

Task and Procedure

In each trial of the delayed-estimation task, participants had to memorize two sequentially presented items (0.4s, inter-stimulus interval 1s), either two visual orientation stimuli, two auditory pitch stimuli or one of each. Crucially, this means that overall working memory load was constant at two items, but unisensory visual load varied between one or two items (see Fig.1). We focused all our analyses on decoding visual stimuli but included the auditory-auditory condition (auditory Load 2) to balance the design behaviourally. Whenever a sample of a given modality was presented, the other modality was masked. After an extended delay (13.8 s), a visual retro-cue ('1' or '2' for 1.2s) indicated whether the first or second item had to be recalled. During the continuous recall phase (4.8s), participants had to adjust a probe orientation or pitch to match the target item by pressing the inner left and right button, and after confirming their response the fixation dot turned green. The next trial started after a brief period of fixation (1.6, 2.4 or 3.2s).

The experiment consisted of 192 trials, divided in four runs with short breaks in between. The item modality was counterbalanced across cued and non-cued items and 12 orientations and pitches were counterbalanced for each modality respectively (see Stimuli). We randomized the experimental conditions within each scanner run to minimize effects of biased training sets for the cross-validation procedure (see fMRI Analyses). Before the main experiment, participants read instructions presented inside the scanner and completed 12 training trials during an anatomical scan. During these training trials, participants received feedback about their performance as the difference between their response and the target item.

Stimuli

Stimuli were generated with MATLAB and Psychtoolbox-3 [43]. The background color was set to grey (RGB [128,128,128]). The fixation dot (radius = 0.1°) was centered on the screen. Orientation stimuli and masks were presented as annuli with a total diameter of 7°, an inner diameter of 1.5°. The visual stimuli consisted of 12 orientations with a 15° distance, from 7.5° to 172.5°, and excluded the cardinal axes, with gaussian blur (gaussian kernel = 1°, sd = 0.5°) on all edges. The spatial frequency was set to one cycle per degree. Visual masks were generated by filtering white noise to only include spatial frequencies between 1 and 4 cycles per degree (code from [17]). During the recall, participants could select their response out of orientations between 1° to 180°, with a precision of 1°. For the auditory stimuli we used 12 pure tones with a distance of one semitone, between 251.63Hz and 493.88Hz, with a logarithmic ramp up and down of 0.02 seconds. White noise was chosen as auditory masks. The volume was adjusted for participants comfort level. During the recall period, a random pitch was played, and participants could select a response between 246.93Hz and 523.24Hz, with a precision of $\textit{Frequency} \times 2^{\pm 1/(70/1)}$. During the recall phase, a longer button press increased the response step size, making it easier for participants to select their response.

Magnetic resonance imaging

Imaging data was recorded on two 3-Tesla Siemens Prisma MRI scanners with a 64 channel head coil. The first 40 participants were scanned at the Berlin Center for Advanced Neuroscience (Charité Berlin, Germany) with the Syngo MR VE11C operating system, a viewing distance of 148cm and screen of dimensions of 52 x 39cm. The subsequent 41 participants at the Center for Cognitive Neuroscience Berlin (Freie Universität Berlin, Germany) with the Syngo MR XA30 operating system, a viewing distance of 106cm and screen of dimensions of 48 x 27.2cm. Stimulus size was adjusted according to the screen-size and viewing distance, so that all participants received the same visual input. The decoding results for both scanners did not differ qualitatively from the combined results. At the beginning of each session, we recorded T1-weighted MPRAGE structural image (208 sagittal slices, TR = 2400 ms, TE = 2.22ms, flip angle = 8°, voxel size = 0.8mm^2 isotropic, field of view = 240*256 mm). For each of the four runs, we recorded functional images with a multi-band accelerated echo planar imaging sequence, with a multi-band factor of 8 (TR = 800ms, TE = 37ms, flip angle = 52°, voxel size = 2 mm^2 isotropic, 72 slices, 2mm inter-slice gap, field of view = 1872mm).

Preprocessing

All preprocessing steps and analyses were performed in MATLAB, using SPM12 [44], The Decoding Toolbox [45], and in-house software. First, we converted the acquired DICOM images into NIfTI format. The intensity-normalized anatomical image (T1) was coregistered to the first functional image of the first block for each participant. During segmentation, we estimated the normalization parameters to the Montreal Neurological Institute (MNI) standard space. Functional images were spatially realigned to the first image of the first block and resliced. To remove slow signal drifts accumulating across each block, we applied a cubic spline interpolation to each voxel's time series ($n_{nodes} = 24$). Afterwards, we applied temporal smoothing by running a moving average of 5 TRs to the data of each block. We used the detrending algorithm and moving average as described in [13].

Identifying ROIs

We focused on the distribution of representations across three regions of interests (ROI) using probabilistic maps of retinotopic regions of the human neocortex [46]. To generate a mask for visual cortex, we combined maps for areas V1 to V3 in standard space. To obtain the bilateral IPS mask, we combined IPS1 to IPS6 from the same probabilistic atlas, and similarly, for the sPCS left and right hemispheres were combined into one mask. After transforming these maps into single-participant space by applying the inverse normalization parameters estimated during preprocessing, we excluded voxels that had less than a 10% probability to belong to each ROI respectively. Secondly, we obtained activity driven maps by estimating a participant-level univariate GLMs for onsets of any orientation presented. After contrasting visual activation and baseline, we selected the 1000 most active voxels within the participant-level anatomical ROIs. We selected the same number of voxels for each ROI in order to make the ROIs as comparable as possible.

fMRI Analyses: pSVR

We analyzed fMRI data using support vector regression (SVR) [47] to continuously reconstruct memorized orientations throughout the delay period. In particular, we projected the angular orientation labels ($\boldsymbol{\theta}$) into sine and cosine labels. This provided a linear measure of these orientations and we predicted each set of these new labels from the multivariate voxel pattern of each ROI with a SVR with a radial basis function kernel (periodic SVR, [13]). We reconstructed the orientation

by using the four-quadrant inverse tangent

$$\hat{\theta} = atan2(\widehat{sin(\theta)}, \widehat{cos(\theta)}).$$

The decoding accuracy scales the difference between presented and predicted orientation label into percent, so that a decoding accuracy of 50% refers to an decoding error of 45°, which can be considered as chance level:

$$DecodingAccuracy = 100 - \frac{|\theta - \hat{\theta}|_{circular}}{\frac{pi}{2}} \times 100.$$

All pSVR analyses were run for each TR individually and cross-validated across runs. In order to increase computational power, we combined all orientations within each of the load conditions (see Fig. 2A). To investigate effects of serial order, we separated the first and second items after the decoding. For the temporal generalization analysis (Fig.3), we trained the pSVR on each time-point and tested on each time-point, both within and across load conditions.

Statistical analyses

We analysed the behavioural data by calculating precision ($\frac{1}{sd}$) for the angular difference between target orientation and response orientation for each condition and participant. We calculated a repeated-measured ANOVA, with load and item position as dependent variables. We then compared the average of each visual load condition, and within each of these conditions of the first and second item position, with paired two-sided t-tests.

To test for significant differences in decoding accuracy between the load conditions (visual Load 1 and 2) and serial order conditions (position 1 and 2), we used paired two-sided t-tests. In particular, we tested decoding accuracy during a period early in the delay, where we assume to find perceptual activation mixed with memory representations (due to hemodynamic lag), and a period later in the delay, where we assume to find information about mnemonic representations without a substantial influence of perceptual representations. We averaged decoding accuracy per participant in each of these two sections of the delay period (early delay: 7.2s-11.2s; late delay: 12s-16s). Then we tested for the difference between the load conditions within each delay period with paired two-sided t-tests and within each condition for decoding accuracy above chance (50%) with two-sided t-tests. We focused all our other analyses on the late delay. Where applicable, for example when comparing results across ROIs or order and load conditions, we corrected for multiple comparisons with Bonferroni correction. For descriptive purposes, we also compared

decoding accuracy to chance level and between conditions (Fig.2B) with two-sided t-tests for each time-point (22 TRs), from first stimulus onset to recall onset. Both analyses were validated with leave one run out cross-validation across participants.

We applied a cluster-based test [48] for the temporal generalization analyses. First, we generated a null distribution by randomly assigning the sign of the above-chance decoding accuracy matrix for each time-point, over 1000 iterations. Within the permutation test, the summed t-value for the largest cluster within null distribution was calculated and compared to our empirical data with a threshold of $p < 0.05$.

Bugs

For the first 11 participants, the number of images recorded in each run was set to 1448 instead of 1460, hence ending immediately after the last trial. Nineteen participants had the same randomisation for the trial order due to a glitch in the code. The results did not change significantly after removing these participants from the analyses.

Acknowledgements

We appreciate Andreea Gui, Damla Cifci, Joana Seabra and Zhiqi Kang's support with the fMRI data collection. We thank Alessandra Souza, Christoph Bledowski, Karolis Degutis, Philipp Deutsch and Rosanne Rademaker for their helpful input in discussions about the experimental design and data analysis. VC and TC were supported by a DFG Emmy Noether Research Group Grant CH 1674/2-1 to TC. SW was funded by the DFG Research Training Group 2386.

Data Availability

Imaging and behavioural data will be shared upon request.

Code Availability

The analysis code will be made available on `github.com/vchopurian/Chopurian_et_al_DistinctRoles`.

Author Contributions

VC and TC designed the study. VC collected the data. VC and TC performed data analysis. SW contributed software and materials. VC and TC wrote the manuscript.

Competing interests

The authors declare no competing interests.

References

[1] Luck, S. J. & Vogel, E. K. The capacity of visual working memory for features and conjunctions. *Nature* **390**, 279–281 (1997).

[2] Gorgoraptis, N., Catalao, R. F., Bays, P. M. & Husain, M. Dynamic updating of working memory resources for visual objects. *Journal of Neuroscience* **31**, 8502–8511 (2011).

[3] Ma, W. J., Husain, M. & Bays, P. M. Changing concepts of working memory. *Nature neuroscience* **17**, 347–356 (2014).

[4] 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. *Journal of Neuroscience* **33**, 6516–6523 (2013).

[5] Sprague, T. C., Ester, E. F. & Serences, J. T. Reconstructions of information in visual spatial working memory degrade with memory load. *Current Biology* **24**, 2174–2180 (2014).

[6] Sutterer, D. W., Foster, J. J., Adam, K. C., Vogel, E. K. & Awh, E. Item-specific delay activity demonstrates concurrent storage of multiple active neural representations in working memory. *PLoS biology* **17**, e3000239 (2019).

[7] Fuster, J. M. Distributed memory for both short and long term. *Neurobiology of learning and memory* **70**, 268–274 (1998).

[8] D'Esposito, M. From cognitive to neural models of working memory. *Philosophical Transactions of the Royal Society B: Biological Sciences* **362**, 761–772 (2007).

[9] 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 (2015).

[10] Christophel, T. B., Klink, P. C., Spitzer, B., Roelfsema, P. R. & Haynes, J.-D. The distributed nature of working memory. *Trends in cognitive sciences* **21**, 111–124 (2017).

[11] Ester, E. F., Anderson, D. E., Serences, J. T. & Awh, E. A neural measure of precision in visual working memory. *Journal of cognitive neuroscience* **25**, 754–761 (2013).

[12] Hallenbeck, G. E., Sprague, T. C., Rahmati, M., Sreenivasan, K. K. & Curtis, C. E. Working memory representations in visual cortex mediate distraction effects. *Nature Communications* **12**, 4714 (2021).

[13] Weber, S., Christophel, T., Görgen, K., Soch, J. & Haynes, J.-D. Working memory and imagery in early visual cortex. *bioRxiv* 2023–02 (2023).

[14] Serences, J. T., Ester, E. F., Vogel, E. K. & Awh, E. Stimulus-specific delay activity in human primary visual cortex. *Psychological science* **20**, 207–214 (2009).

[15] Harrison, S. A. & Tong, F. Decoding reveals the contents of visual working memory in early visual areas. *Nature* **458**, 632–635 (2009).

[16] Lorenc, E. S., Sreenivasan, K. K., Nee, D. E., Vandenbroucke, A. R. & D'Esposito, M. Flexible coding of visual working memory representations during distraction. *Journal of Neuroscience* **38**, 5267–5276 (2018).

[17] Rademaker, R. L., Chunharas, C. & Serences, J. T. Coexisting representations of sensory and mnemonic information in human visual cortex. *Nature neuroscience* **22**, 1336–1344 (2019).

[18] Fuster, J. M. & Jervey, J. P. Inferotemporal neurons distinguish and retain behaviorally relevant features of visual stimuli. *Science* **212**, 952–955 (1981).

[19] Funahashi, S., Bruce, C. J. & Goldman-Rakic, P. S. Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *Journal of neurophysiology* **61**, 331–349 (1989).

[20] Goldman-Rakic, P. S. Cellular basis of working memory. *Neuron* **14**, 477–485 (1995).

[21] Bettencourt, K. C. & Xu, Y. Decoding the content of visual short-term memory under distraction in occipital and parietal areas. *Nature neuroscience* **19**, 150–157 (2016).

[22] Constantinidis, C., Franowicz, M. N. & Goldman-Rakic, P. S. The sensory nature of mnemonic representation in the primate prefrontal cortex. *Nature neuroscience* **4**, 311–316 (2001).

[23] Wimmer, K., Nykamp, D. Q., Constantinidis, C. & Compte, A. Bump attractor dynamics in prefrontal cortex explains behavioral precision in spatial working memory. *Nature neuroscience* **17**, 431–439 (2014).

[24] Spitzer, B. & Blankenburg, F. Supramodal parametric working memory processing in humans. *Journal of Neuroscience* **32**, 3287–3295 (2012).

[25] Leavitt, M. L., Pieper, F., Sachs, A. J. & Martinez-Trujillo, J. C. A quadrantic bias in prefrontal representation of visual-mnemonic space. *Cerebral Cortex* **28**, 2405–2421 (2018).

[26] Yan, C., Christophel, T. B., Allefeld, C. & Haynes, J.-D. Categorical working memory codes in human visual cortex. *NeuroImage* **274**, 120149 (2023).

[27] Franconeri, S. L., Alvarez, G. A. & Cavanagh, P. Flexible cognitive resources: competitive content maps for attention and memory. *Trends in cognitive sciences* **17**, 134–141 (2013).

[28] Cohen, M. A., Konkle, T., Rhee, J. Y., Nakayama, K. & Alvarez, G. A. Processing multiple visual objects is limited by overlap in neural channels. *Proceedings of the National Academy of Sciences* **111**, 8955–8960 (2014).

[29] Baddeley, A. Working memory: looking back and looking forward. *Nature reviews neuroscience* **4**, 829–839 (2003).

[30] Kiyonaga, A. & D'Esposito, M. *Competition and Control during Working Memory* (Cambridge University Press, 2020).

[31] Wing, E. A., Burles, F., Ryan, J. D. & Gilboa, A. The structure of prior knowledge enhances memory in experts by reducing interference. *Proceedings of the National Academy of Sciences* **119**, e2204172119 (2022).

[32] Overkott, C. & Souza, A. S. Verbal descriptions improve visual working memory but have limited impact on visual long-term memory. *Journal of Experimental Psychology: General* **151**, 321 (2022).

[33] Chunharas, C., Hettwer, M. D., Wolff, M. J. & Rademaker, R. L. A gradual transition from veridical to categorical representations along the visual hierarchy during working memory, but not perception. *bioRxiv* 2023–05 (2023).

[34] Inoue, M. & Mikami, A. Prefrontal activity during serial probe reproduction task: encoding, mnemonic, and retrieval processes. *Journal of neurophysiology* **95**, 1008–1041 (2006).

[35] Manohar, S. G., Pertzov, Y. & Husain, M. Short-term memory for spatial, sequential and duration information. *Current opinion in behavioral sciences* **17**, 20–26 (2017).

[36] Kalm, K. & Norris, D. A shared representation of order between encoding and recognition in visual short-term memory. *NeuroImage* **155**, 138–146 (2017).

[37] Xie, Y. *et al.* Geometry of sequence working memory in macaque prefrontal cortex. *Science* **375**, 632–639 (2022).

[38] Broadbent, D. E. & Broadbent, M. H. Recency effects in visual memory. *The Quarterly Journal of Experimental Psychology Section A* **33**, 1–15 (1981).

[39] Fuster, J. M. The cognit: a network model of cortical representation. *International Journal of Psychophysiology* **60**, 125–132 (2006).

[40] Yu, Q. & Shim, W. M. Temporal-order-based attentional priority modulates mnemonic representations in parietal and frontal cortices. *Cerebral Cortex* **29**, 3182–3192 (2019).

[41] Botvinick, M. & Watanabe, T. From numerosity to ordinal rank: a gain-field model of serial order representation in cortical working memory. *Journal of Neuroscience* **27**, 8636–8642 (2007).

[42] World Medical Association. World medical association declaration of helsinki: ethical principles for medical research involving human subjects. *Jama* **310**, 2191–2194 (2013).

[43] Brainard, D. H. & Vision, S. The psychophysics toolbox. *Spatial vision* **10**, 433–436 (1997).

[44] Ashburner, J. *et al.* Spm12 manual. *Wellcome Trust Centre for Neuroimaging, London, UK* **2464** (2014).

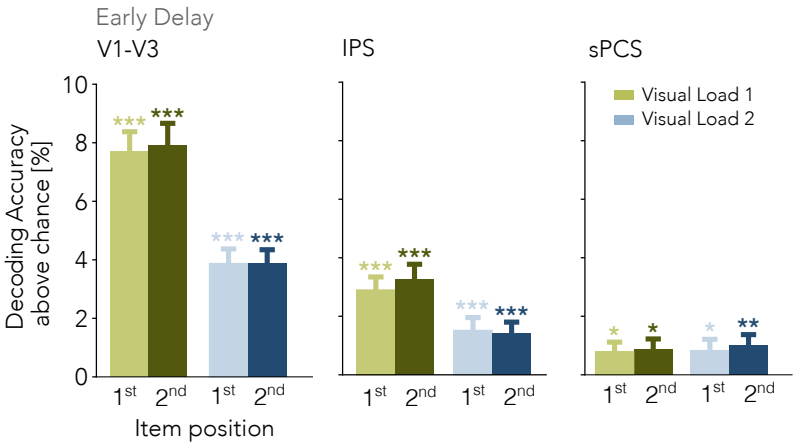
[45] Hebart, M. N., Görgen, K. & Haynes, J.-D. The decoding toolbox (tdt): a versatile software package for multivariate analyses of functional imaging data. *Frontiers in neuroinformatics* **8**, 88 (2015).

[46] Wang, L., Mruczek, R. E., Arcaro, M. J. & Kastner, S. Probabilistic maps of visual topography in human cortex. *Cerebral cortex* **25**, 3911–3931 (2015).

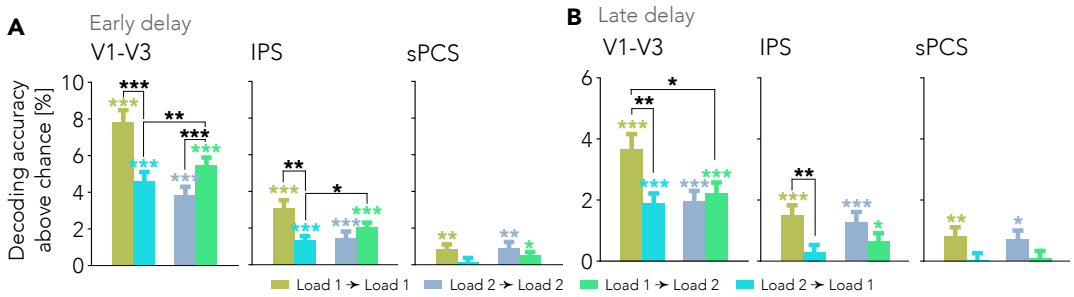
[47] Chang, C.-C. & Lin, C.-J. Libsvm: a library for support vector machines. *ACM transactions on intelligent systems and technology (TIST)* **2**, 1–27 (2011).

[48] Maris, E. & Oostenveld, R. Nonparametric statistical testing of eeg-and meg-data. *Journal of neuroscience methods* **164**, 177–190 (2007).

Supplementary Material



Supplementary Fig. 1: Early delay serial order results. Decoding of first and second item during the early delay period. In none of our ROIs did we find a difference between the first and second orientation. (V1-V3 visual Load 1: $t_{(80)} = -0.47, p = 1.0$, V1-V3 visual Load 2: $t_{(80)} = 0.025, p = 1.0$; IPS visual Load 1: $t_{(80)} = -0.94, p = 1.0$, IPS visual Load 2: $t_{(80)} = 0.22, p = 1.0$; sPCS visual Load 1: $t_{(80)} = -0.17, p = 1.0$, sPCS visual Load 2: $t_{(80)} = 0.46, p = 1.0$; all p-values are Bonferroni corrected). $^{*}p < .05$, $^{**}p < .01$ or $^{***}p < .001$. Error bars show the SEM.



Supplementary Fig. 2: Condition cross-decoding results. Cross-condition decoding. We trained on each visual load condition and tested on each visual load condition. Statistical significance is only shown for comparisons with cross-decoding results to reduce information in plots. These bars correspond to the diagonal in Fig3C and D. **A** Above chance test results for early delay period: V1-V3, Load 1-Load 2: $t_{(80)} = 13.52, p < 0.001$; V1-V3, Load 2-Load 1: $t_{(80)} = 9.90, p < 0.001$; IPS, Load 1-Load 2: $t_{(80)} = 7.27, p < 0.001$; IPS, Load 2-Load 1: $t_{(80)} = 5.49, p < 0.001$; sPCS, Load 1-Load 2: $t_{(80)} = 2.55, p = 0.01$; sPCS, Load 2-Load 1: $p > 0.43$. **B** Above chance test results for late delay period: V1-V3, Load 1-Load 2: $t_{(80)} = 6.36, p < 0.001$; V1-V3, Load 2-Load 1: $t_{(80)} = 5.81, p < 0.001$; IPS, Load 1-Load 2: $t_{(80)} = 2.56, p = 0.01$; IPS, Load 2-Load 1: $p > 0.2$; sPCS, Load 1-Load 2: $p > 0.6$; sPCS, Load 2-Load 1: $p > 0.9$.
Asterisks indicate significance levels of $*p < .05$, $**p < .01$ or $***p < .001$. Error bars show the SEM. $*p < .05$, $**p < .01$ or $***p < .001$. Error bars show the SEM.